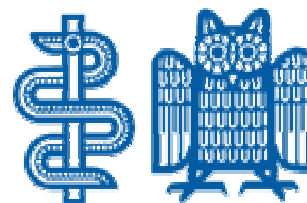


Aus dem Bereich Klinische Medizin  
der medizinischen Fakultät  
der Universität des Saarlandes, Homburg/Saar



## **Verbreitung der humanen endogenen Retroviren HERV-K113 und HERV-K115 in der menschlichen Population**

*Dissertation zur Erlangung des Grades eines Doktors der Medizin*

**der Medizinischen Fakultät**  
der UNIVERSITÄT DES SAARLANDES

2008

vorgelegt von: Johannes Klatt  
geb. am: 10.02.1981 in Greifswald

---

Aus dem Institut für Virologie  
Universitätsklinikum des Saarlandes, Homburg/Saar  
(Direktor: Prof. Dr. Nikolaus Müller-Lantzsch)

---

Meinen Eltern

## Inhaltsverzeichnis

Tabellenverzeichnis	VI
Abbildungsverzeichnis	VII
Abkürzungsverzeichnis	X
<b>1. Zusammenfassung .....</b>	<b>1</b>
1.1 Zusammenfassung .....	1
1.2 Summary .....	2
<b>2. Einleitung.....</b>	<b>4</b>
2.1 Viren .....	4
2.2 Retroviren .....	4
2.2.1 Taxonomie .....	5
2.2.2 Morphologie .....	6
2.2.3 Retrovirales Genom .....	7
2.2.4 Infektionsverlauf und Replikationszyklus.....	9
2.2.5 Auswirkungen retroviraler Integration.....	11
2.3 Endogene Retroviren .....	12
2.3.1 Humane endogene Retroviren .....	13
2.3.2 Klassifizierung.....	14
2.3.3 Das humane endogene Retrovirus K.....	16
2.3.4 Strukturelle Variabilität innerhalb der HERV-K-Familie.....	19
2.3.5 Einfluss von HERV-K auf den Wirtsorganismus .....	20
2.3.6 Die retroviralen Elemente HERV-K113 und HERV-K115 .....	21
2.4 Ziele der Arbeit .....	22
<b>3. Material und Methodik.....</b>	<b>24</b>
3.1 DNA-Isolierung.....	24
3.1.1 Isolierung über Anionenaustauschersäulen.....	24
3.1.2 Isolierung hochmolekularer DNA aus Blut durch Salz-Chloroform-Extraktion.....	25
3.2 Photometrische Konzentrationsbestimmung .....	26
3.3 Polymerase-Kettenreaktion.....	27
3.3.1 Reaktionsbedingungen .....	28
3.3.2 Die verwendeten Oligonukleotidprimer.....	29

3.3.3	Qualitätskontrolle der isolierten DNA .....	30
3.3.4	Nested-PCR .....	31
3.4	Gelelektrophorese.....	31
3.4.1	Agarose-Gelelektrophorese .....	32
3.4.2	Polyacrylamid-Gelelektrophorese .....	33
3.5	Reisolierung von DNA-Fragmenten aus Agarosegelen .....	34
3.6	Spaltung von DNA durch Restriktionsendonukleasen .....	35
3.7	Vermehrung von PCR-Fragmenten durch Bakterien .....	38
3.7.1	Klonierung.....	38
3.7.2	Transformation .....	40
3.7.3	Screening positiver Klone .....	41
3.8	Präparation von Plasmid-DNA in kleinem Maßstab .....	43
3.9	Kontrollverdau .....	43
3.10	Präparation von Plasmid-DNA in großem Maßstab .....	45
3.11	Sequenzierung von DNA-Fragmenten .....	47
<b>4.</b>	<b>Ergebnisse .....</b>	<b>48</b>
4.1	Zusammensetzung der untersuchten Population.....	48
4.2	Die verwendeten PCR-Schemata .....	48
4.2.1	Nachweis von HERV-K113 .....	49
4.2.2	Analyse der Präintegrationsstelle und der duplizierten Region .....	50
4.2.3	Nachweis von HERV-K115 .....	52
4.3	Verteilung der retroviralen Elemente HERV-K113 und HERV-K115 .....	54
4.3.1	Vergleich der Frequenz innerhalb der untersuchten Gruppe.....	55
4.4	Der Nachweis von Homozygotie bei dem HERV-K113-Provirus .....	56
4.4.1	Theoretische Vorüberlegung .....	56
4.4.2	Untersuchung der Homo- und Heterozygotie mit Hilfe von Restriktionsendonukleasen.....	57
4.4.3	Untersuchung von heterozygoten Merkmalsträgern.....	60
4.4.4	Untersuchung der Merkmalsträger mit <i>SspI</i> .....	61
4.4.5	Untersuchung der Merkmalsträger mit <i>AflIII</i> .....	62
4.5	Sequenzanalyse der duplizierten Region und der Präintegrationsstelle .....	63
4.6	Sequenzanalyse des <i>env</i> - und des <i>gag</i> -Gens.....	63
4.6.1	Das <i>env</i> -Gen .....	63
4.6.2	Das <i>gag</i> -Gen.....	64

Inhaltsverzeichnis.....	V
-------------------------	---

<b>5. Diskussion .....</b>	<b>65</b>
5.1 Fragestellungen der Arbeit.....	65
5.2 Genfrequenz und Bedeutung von HERV-K113 und HERV-K115 für den Menschen .....	66
5.3 Homozygotie des humanen endogenen Retrovirus K113 .....	70
5.4 HERV-K 113 und seine provirale Integrität .....	71
<b>6. Literaturverzeichnis .....</b>	<b>74</b>
<b>7. Anhang.....</b>	<b>84</b>
7.1 Die Sequenz der duplizierte Region.....	84
7.2 Das sequenzierte <i>env</i> -Gen.....	84
7.3 Das sequenzierte <i>gag</i> -Gen .....	92
<b>8. Danksagung .....</b>	<b>99</b>
<b>9. Lebenslauf.....</b>	<b>101</b>

Tabelle 1:	Klassifikation der Retroviren nach ICTV, Stand 2005
Tabelle 2:	Klassifikation humaner endogener Retroviren
Tabelle 3:	Pipettierschema für den PCR-Ansatz
Tabelle 4:	Ablauf des PCR-Programms JM3min mit einer auf die verwendeten Oligonukleotidprimer abgestimmten Annealing-Temperatur (TM)
Tabelle 5:	verwendete Oligonukleotidprimer mit Annealingtemperatur und Accession-Nr.
Tabelle 6:	benutzte Enzyme mit Schnittstelle, passendem Puffer und eingesetzter Konzentration
Tabelle 7:	Pipettierschema für Ligationsansatz
Tabelle.8:	Pipettierschema für den Orientierungsverdau mit <i>SacI</i> , <i>SacII</i> und <i>NotI</i> und <i>SphI</i> und den jeweils benötigten Agenzien
Tabelle 9:	Verwendete Puffer und deren Zusammensetzung
Tabelle 10:	nachgewiesene Häufigkeit bzw. Genfrequenz
Tabelle 11:	Häufigkeitsverteilung von HERV-K113 und HERV-K115 in den untersuchten Gruppen
Tabelle 12:	Ergebnis der Sequenzanalyse des <i>env</i> -Leserahmens
Tabelle 13:	Ergebnis der Sequenzanalyse des <i>gag</i> -Leserahmens

- Abb.1: schematische Darstellung eines retroviralen Virions am Beispiel des HI-Virus
- Abb.2: Schematische Darstellung der retroviralen Erbinformation in Form einer mRNA mit 5'Cap, Polyadenylierung am 3'-Ende und Primerbindungsstelle (PBS) für die *reverse Transkriptase* (RT)
- Abb.3: schematische Darstellung des proviralen Aufbaus von HERV-K
- Abb.4: a) Typ 1 Provirus mit Deletion ( $\Delta$ ) und Np9 als Spleißprodukt  
b) Typ 2 Provirus mit Spleißdonor (SD) und-akzeptor (SA) für Rec
- Abb.5: Proviraer Aufbau von HERV-K113 und HERV-K115
- Abb.6: Phasen der Polymerase-Kettenreaktion
- Abb.7: 191bp umfassendes PCR-Produkt nach Reaktion mit den Primern GAPDH N / GAPDH M auf 3%Agarosegel
- Abb.8: Strukturformel von Ethidiumbromid
- Abb.9: Prinzip der Reisolierung von DNS-Fragmenten aus Agarosegelen
- Abb.10: pGEM®-T Vektor
- Abb.11: schematische Darstellung der Klonierung
- Abb.12: Anzucht transformierter Zellen
- Abb.13: a) 5-Brom-4-chlor-3-indoxyl- $\beta$ -D-galactopyranosid  
b) Isopropyl- $\beta$ -D-thiogalactopyranosid
- Abb.14: Schematische Darstellung des Blue-White-Screenings

- Abb.15: multiple cloning site des pGEM-Vektors mit den dazugehörigen Enzymschnittstellen
- Abb.16: Beispiel eines Kontrollverdaus
- Abb.17: Verdau der Plasmid-DNA durch die Restriktionsendonukleasen
- Abb.18: Fragmente nach der PCR mit den Primern K113 for / LTRgag rev
- Abb.19: Fragmente nach der PCR mit den Primern env8455 for / K113 rev
- Abb.20: PCR mit den Oligonukleotidprimern K113 for / K113 rev und K113DuF2 / K113 rev, der Präintegrationsstelle und der duplizierten Region
- Abb.21: PCR mit den Primern K115r2 / LTRgag rev
- Abb.22: PCR mit den Primern env8455 for / K115f2
- Abb.23: PCR-Schema zum Nachweis von Homozygotie mit den Primern K113 for und K113 rev
- Abb.24: Schematische Darstellung der Restriktionsschnittstellen in den Fragmenten nach PCR mit den Primern K113 for / K113 rev für die Restriktionsendonukleasen *SspI* in der Präintegrationsstelle und *AfIII* in der duplizierten Region
- Abb.25: Zu erwartende Banden auf einem 15%Polyacrylamidgel nach Einsatz der Restriktionsendonukleasen *SspI* und *AfIII*
- Abb.26: a) ungeschnittenes PCR-Fragment von vier heterozygoten Merkmalsträgern auf einem 15% Polyacrylamidgel nach Verdau mit *AfIII*; b) zu erwartendes Ergebnis nach Verdau des PCR-Produktes mit *AfIII*, bei dem die duplizierte Region nicht aber die Präintegrationsstelle geschnitten wurde



Abb.27: Nachweis von homozygoten und heterozygoten Merkmalsträgern durch Verdau der mittels PCR gewonnenen Fragmente mit der Restriktionsendonuklease *SspI*

Abb.28: Nachweis von homozygoten und heterozygoten Merkmalsträgern durch Verdau der mittels PCR gewonnenen Fragmente mit der Restriktionsendonuklease *AflIII*

A	Adenin
Abb	Abbildung
AIDS	Acquired Immunodeficiency Virus
b	Base
bp	Basenpaar
C	Cytosin
Ca	Karzinom
cORF	central Open Reading Frame
DNA	Desoxyribonukleinsäure
dNTP	Desoxyribonukleosidtriphosphat
ds	double strand = Doppelstrang
EDTA	Ethylendiamintetraacetat
env	Envelope-Gen
G	Guanin
gag	gruppenspezifisches Antigen
HIV	humanes Immundefizienzvirus
HTLV	humanes T-Zelllymphotropes Virus
LTR	Long Terminal Repeat
M	molar
MLV	Mouse Leukaemia Virus
mM	millimolar
MMTV	Mouse Mammarian Tumor Virus
mRNA	messenger RNA
ng	Nanogramm = $10^{-9}$ g
nm	Nanometer = $10^{-9}$ m
PBS	Primer binding site
PCR	Polymerasekettenreaktion
pol	Polymerase-Gen
pr / pro	Protease-Gen
RNA	Ribonukleinsäure
RSV	Rous Sarcoma Virus
SNV	Splen Nekrosis Virus
SNP	Single Nucleotide Polymorphism
ss	single strand = Einzelstrang

T	Thymin
TAE	Tris-Acetat-EDTA
T <sub>m</sub>	Annealing-, Schmelztemperatur
Tris	Triaminomethan
tRNA	Transfer-RNA
U	Unit
U/min	Umdrehungen pro Minute
UV	Ultraviolett

Die Erklärungen der hier nicht aufgeführten Abkürzungen sind dem Text zu entnehmen.

# 1. Zusammenfassung

## 1.1 Zusammenfassung

Die Integration exogener Retroviren in Keimbahnzellen und die sich daran anschließende vertikale Weitergabe von Generation zu Generation vor Millionen Jahren sind die Gründe, warum heute bis zu 8% des humanen Genoms aus endogenen retroviralen Elementen bestehen. Durch die Entschlüsselung der menschlichen Erbinformation konnten viele dieser Elemente verschiedenen Familien zugeordnet werden. Die HERV-K-Familie nimmt unter ihnen eine Sonderstellung ein. Zum einen sind einige ihrer Mitglieder zur Transkription und Partikelbildung fähig. Zum anderen wird HERV-K mit der Entstehung von Keimbahntumoren in Verbindungen gebracht. Die bezogen auf ihre Integration jüngsten Mitglieder sind die retroviralen Elemente HERV-K113 und HERV-K115. Ihre Integration fand in einem zeitlichen Rahmen von 200000 – 1000000 Jahren statt. Bei beiden handelt es sich um Volllängenproviren mit unversehrten Präintegrationsstellen. Für HERV-K113 konnten intakte offene Leserahmen für alle retroviralen Gene nachgewiesen werden. Damit besitzt dieses endogene Retrovirus alle Voraussetzungen um erneut Zellen zu infizieren.

Untersuchungen verschiedener Populationen ergaben, dass beide Proviren nicht in jedem menschlichen Genom zu finden sind. Vielmehr zeigen sich bezüglich der Prävalenz erhebliche Schwankungen beruhend auf Herkunft und Abstammung der untersuchten Individuen. Dieser Insertionspolymorphismus ist die Grundlage für die vorliegende Arbeit. Mit Hilfe der Polymerasekettenreaktion wurde in einer 207 Individuen umfassenden Population kaukasischen Ursprungs für HERV-K113 eine Häufigkeit von 13,5% und für HERV-K115 eine Häufigkeit von 4,4% gefunden. Darüber hinaus vergleicht diese Arbeit das Vorkommen beider Proviren in einer Gruppe an Seminomen Erkrankter (n=28) mit einer gesunden Kontrollgruppe (n=80). Dabei konnte festgestellt werden, dass sich mit 11,1% zu 18,8% für HERV-K113 und 7,4% zu 7,5% für HERV-K115 die Prävalenzen in beiden Populationen nicht signifikant unterscheiden.

Die Suche nach homozygoten Merkmalsträgern für HERV-K113 stellt einen weiteren Aspekt dieser Arbeit dar. Der Einsatz der Restriktionsendonukleasen *Afl*III und *Ssp*I ermöglichte, innerhalb einer Population, die sich aus 46 Individuen verschiedener Ethnitäten

zusammensetzte, drei reinerbige Merkmalsträger zu finden. Dabei konnte je einem dieser Individuen eine afrikanische und eine asiatische Herkunft zugeordnet werden. Die dritte Probe stammt von einem Aborigines.

In der sich daran anschließenden Sequenzanalyse der *env*- und *gag*-Gene von HERV-K113 wurden interindividuelle Unterschiede in der Integrität des proviralen Genoms untersucht. Zu dem wurde geprüft, ob etwaige Polymorphismen zufällig verteilt oder auf bestimmte Mutationen selektiert werden. Bei den sieben *env*- und fünf *gag*-Genen, die untersucht wurden, konnten keine signifikanten Unterschiede gefunden und damit Sequenzhomologie für beide Gene nachgewiesen werden.

Die in der vorliegenden Arbeit erhobenen Daten konnten die Hypothese bestätigen, dass sich die Verteilung der humanen endogenen Retroviren HERV-K113 und HERV-K115 im menschlichen Genom nicht global bestimmen lässt, sondern die Prävalenz stark von der ethnischen Zusammensetzung der untersuchten Population abhängt. Darüber hinaus scheint es keinen direkten Zusammenhang zwischen der Integration beider Proviren und der Entstehung von Keimbahntumoren zu geben.

## 1.2 Summary

The integration of exogenous retroviruses into germline cells and the following vertical pass on from generation to generation millions of years ago are the reasons why today up to 8% of the humane genome consist of endogenous retroviral elements. Due to the decoding of the humane genome many of these elements were able to be attached to different families. Among those groups the HERV-K-Family plays a special role. On the one hand some of their members are able to transcript. On the other hand it is very likely that HERV-K is associated with germcell carcinomas. According to their integration the retroviral elements HERV-K113 and HERV-K115 are the youngest members. This event took place in a period of time about 200.000 – 1.000.000 years ago. Both are full-length proviruses with intact preintegration sites. HERV-K113 also has open reading frames for all retroviral genes. With that this endogenous retrovirus has all opportunities to reinfect new cells even today.

Studies in different populations showed that both proviruses cannot be found in every humane DNA. Based on the origin and ancestry of the analysed individuals considerable fluctuations are to be found with regard to prevalence on the contrary. This insertional polymorphism is

the foundation for the proposed work. With the use of the polymerase chain reaction for HERV-K113 a frequency of 13.5% and for HERV-K115 a frequency of 4.4% was found in a population with 207 individuals of caucasian origin. Beyond that this work compares the distribution of both proviruses in one group containing of patients with seminomas (n=28) to a healthy control group (n=80). Here it was found that with 11.1% to 18.8% for HERV-K113 and 7.4% to 7.5% the prevalences do not differ significantly.

Another aspect of this work is the search for homozygosity in the population positive for HERV-K113. Putting the restriktion endonucleases *Afl*III and *Ssp*I into action it was possible to find homozygosity for three members in a population containing of 46 individuals of different origin. One specimen came from Africa, an other from Asia and the third from an Aborigines.

In the following sequence analysis of the *env* and *gag* genes of HERV-K113 interindividual differences according to the integration of the proviral genome were examined. Besides it was tested whether possible polymorphisms are distributed by chance or selected to special mutations. Among the seven *gag* and five *env* genes that were analysed no significant difference could be found and hence homology of sequences was shown for both genes.

The data risen in the proposed work could confirm the hypothesis that the distribution of the humane endogenous retroviruses HERV-K113 and HERV-K115 in the humane genome cannot be tested for globally but it strongly depends on the ethnic composition of the population analysed. Beyond that there seems to be no correlation between the integration of one of the provirus and the formation of germline carcinomas.

## 2. Einleitung

### 2.1 Viren

Bei der Suche nach dem Erreger der Maul- und Klauenseuche fand Friedrich Loeffler 1897 einen Erreger, der von keinem bakteriendichten Filter zurückgehalten wurde, im Lichtmikroskop nicht sichtbar war und sich nicht anzüchten ließ. Die Tatsache, dass er die Krankheit mit gereinigtem Serum von einem Tier auf ein anderes übertragen konnte, brachte diesem Erreger die Bezeichnung Virus (lat.: schleimige Absonderung; Gift) ein. Diese Partikel bestehen lediglich aus ihrem Genom, das von einer Kapsel umschlossen wird und, abhängig von der Art, einer Hülle. Aus der Abwesenheit von jedweden Zellorganellen wie Mitochondrien oder Ribosomen, die Voraussetzung für einen Zellstoffwechsel sind, ergibt sich eine obligat intrazelluläre Vermehrung. Durch das Einschleusen des viralen Genoms in das der Wirtszelle bedient sich das Virus des zellulären Proteinbiosyntheseapparates für seine eigene Vermehrung. Eine Einteilung nach Art und Polarität des Genoms, nach Form der Kapsel und nach Vorhandensein einer Hülle ordnet jedes Virus einer bestimmten Familie zu. So gehört das Hepatitis-B-Virus beispielsweise zu der Gattung der *Orthohepadnavirinae*, der Familie der *Hepadnaviridae* und diese Familie zu den behüllten Viren mit doppelsträngiger DNA (dsDNA).

### 2.2 Retroviren

Die Familie der *Retroviridae* unterscheidet sich von anderen Virenfamilien durch die Fähigkeit, ihre Erbinformation, die als Einzelstrang-RNA vorliegt, in einen DNA-Doppelstrang umzuschreiben und diesen in das Genom der Wirtszelle zu integrieren. Das zentrale Dogma der Molekularbiologie, dass genetische Information nur unidirektional von der DNA in Richtung RNA weitergeleitet wird, hat damit nicht länger Bestand. Howard Temin und David Baltimore beschrieben unabhängig voneinander 1970 das Enzym *Reverse Transkriptase*, das diesen Schritt katalysiert. Beide erhielten dafür 1975 zusammen mit Renato Dulbecco den Nobelpreis für Medizin bzw. Physiologie (Temin und Mizutani, 1970; Baltimore et al., 1970). Erste Hinweise für das Vorhandensein solcher Partikel gab es

allerdings schon Anfang des 20. Jahrhunderts. So wiesen Vilhelm Ellermann und Oluf Bang 1908 nach, dass es sich bei der Hühnerleukose um eine virale Infektion handelt und nannten den Erreger Avian Leukosis Virus (ALV) (Ellermann und Bang, 1908). Drei Jahre später gelang Peyton Rous die zellfreie Übertragung eines Sarkoms bei Hühnern (Rous, 1911). Dieses Virus, das Rous Sarkoma Virus (RSV), trägt heute noch den Namen seines Entdeckers. Im weiteren Verlauf des Jahrhunderts wurden stetig neue Erreger gefunden, die später den Retroviren zugeordnet werden konnten. Der erste Nachweis eines durch Retroviren hervorgerufenen Tumors bei Säugetieren gelang 1936 John Bittner (Bittner, 1936). Mit der Entdeckung des Mouse Mammary Tumor Virus (MMTV) konnte er zeigen, dass Brustkrebs bei infizierten Mäusen über die Muttermilch übertragbar ist. Die erste Beschreibung dessen, was sich später als Infektion durch Lentiviren herausstellte, wurde 1904 gemacht (Vallée und Carré, 1904). Vallée und Carré berichteten über eine infektiöse Anämie bei Pferden. Der Nachweis des Maedi-Visna-Virus (MVV) bei Schafen bekräftigte das Konzept der Slow-Virusinfektion durch Lentiviren (lat. lentus = langsam). Eine durch derartige Viren hervorgerufene Erkrankung zeichnet sich durch eine lange Inkubationszeit und einen schleichenden Verlauf aus (Sigurdsson, 1954). Der heute bekannteste Vertreter dieser Gattung ist das HI-Virus.

## 2.2.1 Taxonomie

(Fauquet et al., 2005)

Seit 1971 befasst sich das International Committee on Taxonomy of Viruses (ICTV) mit der Klassifizierung von Viren. Sie teilt die Familie der Retroviren anhand morphologischer Besonderheiten in zwei Unterfamilien, die *Orthoretrovirinae* und die *Spumaretrovirinae*, ein. Weitere Unterteilungen innerhalb der Unterfamilien in verschiedene Gattungen ergeben sich aus einer unterschiedlichen Oberflächenbeschaffenheit, Genomsequenz und Sequenz der Genomprodukte, dem Vorhandensein von Onkogenen und dem Wirtsspektrum (Tabelle 1). Ein weiterer Aspekt, nach dem man die Retroviren einteilen kann, ist die Fähigkeit zum alternativen Spleißen. Vertreter dieser Gruppe werden als komplex bezeichnet und sind in der Lage, die Information eines Gens zur Synthese verschiedener Proteine zu nutzen. Das Virus benutzt diese akzessorischen Proteine, um die Kontrolle über die Wirtszelle auszubauen oder enzymatischen Abbauprozessen auszuweichen. Zu ihnen zählen die Lentiviren, die Spumaviren und das Humane T-Zell-lymphotrope Virus. Auch einige humane endogene



Retroviren sind in der Lage ihr Genom alternativ zu spleißen (Löwer et al., 1993; Boese et al., 2000).

Unterfamilie	Gattung	Beispiel
<i>Orthoretrovirinae</i>	Alpharetrovirus	Rous-Sarkom-Virus, Vogel-Leukose-Virus
	Betaretrovirus	Mausbrusttumorvirus,
	Gammaretrovirus	Mäuse-Leukämie-Virus, Katzen-Leukämie-Virus
	Deltaretrovirus	Humanes T-Zell-lymphothrope Virus 1 / 2
	Epsilonretrovirus	Walleye-Dermale-Sarkom-Virus
	Lentivirus	Humanes Immunodefizienz Virus 1 / 2 Maedi-Visna-Virus
<i>Spumaretrovirinae</i>	Spumavirus	Pferde-Schaumvirus, Katzen-Schaumvirus

Tabelle 1: Klassifikation der Retroviren nach ICTV, Stand 2005

## 2.2.2 Morphologie

Die Vertreter der *Retroviridae* sind umhüllte RNA-Viren, deren Virion einen Durchmesser von 80 bis 100nm aufweist. Das Genom liegt mit einer Länge von 7-12 kb als eine einzelsträngige lineare nicht segmentierte Ribonukleinsäure positiver Polarität vor (ss-(+)-RNA). Die Abbildung 1 zeigt repräsentativ für alle Retroviren den Aufbau eines Viruspartikels anhand des HI-Virus. Zusammen mit den an die RNA angelagerten Nukleoproteinen entsteht das Nukleokapsid (NC), das gemeinsam mit den Replikationsenzymen *Reverse Transkriptase* (RT), *Integrase* (IN) und *Protease* (PR) von dem Kapsid (CA) umschlossen wird. Die Virushülle (Env) ist der einzige Bestandteil eines Virions der nicht durch virale Gene kodiert wird, sondern von der Wirtszelle stammt. Allerdings

werden in diese Hülle Transmembranproteine (TM) und Glykoproteine (SU) eingelagert, die für jedes Virus spezifisch und die Infektion weiterer Zellen essentiell sind. Zwischen Virushülle und dem Kapsid befindet sich die Matrix (MA), die ebenfalls Produkt des viralen Genoms ist (Coffin, 1992).

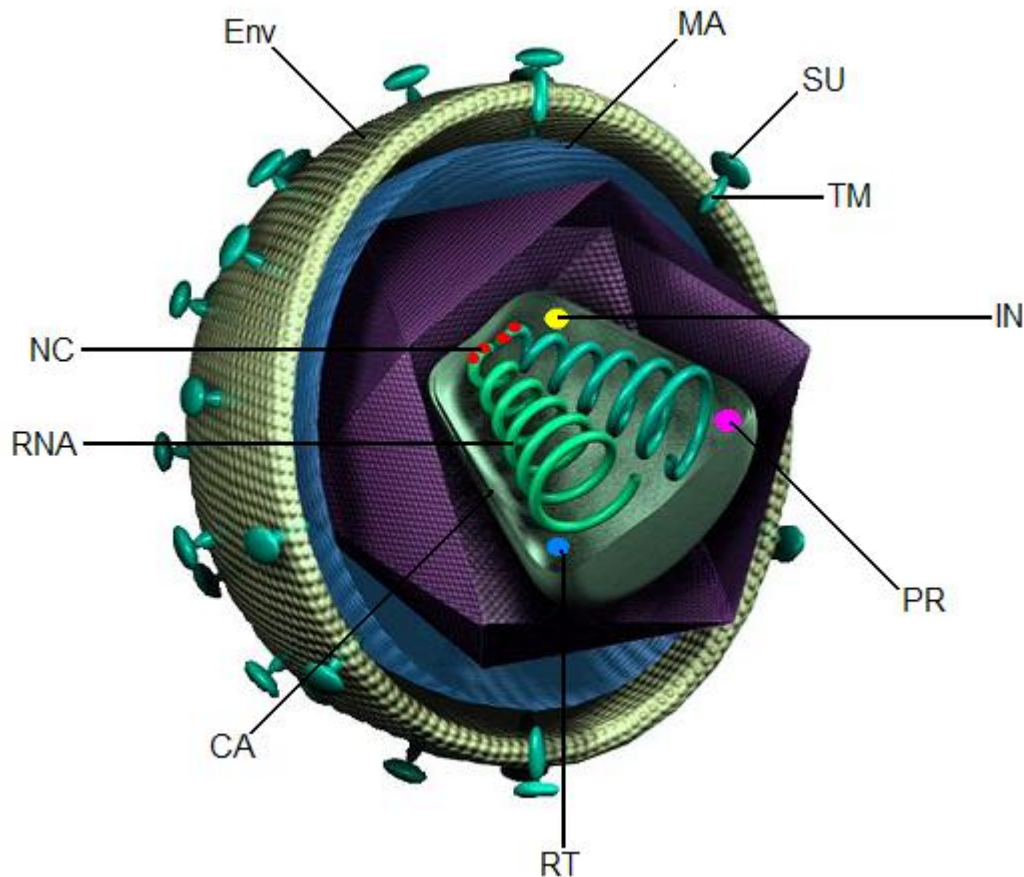


Abb.1: schematische Darstellung eines retroviralen Virions am Beispiel des HI-Virus ( CA = Kapsid, Env = Virushülle, IN = *Integrase*; MA = Matrix; NC = Nukleokapsid; PR = *Protease*; RNA = Virusgenom; RT = *reverse Transkriptase*; SU = Glykoprotein, TM = Transmembranprotein)

### 2.2.3 Retrovirales Genom

(*Retroviruses*, Coffin, Hughes, Varmus 1997)

Die retrovirale Erbinformation liegt im Virion in Form einer Boten-RNA mit einem 5'-Cap und einem PolyA-Teil am 3'-Ende vor (Furuichi et al., 1975; Gillespie et al., 1972). Diese zusätzlichen Sequenzen dienen der Initiation und Regulation der späteren Translation. So werden mRNA-Moleküle mit einem kürzeren Poly-A-Teil beispielsweise schneller translatiert

als solche mit einer längeren Sequenz. Um die einzelnen Abschnitte des Genoms und deren Aufgaben besser beschreiben zu können, wird im Folgenden von dem in die DNA der Wirtszelle integrierten retroviralen Genom, dem Provirus, ausgegangen (Abb. 2).

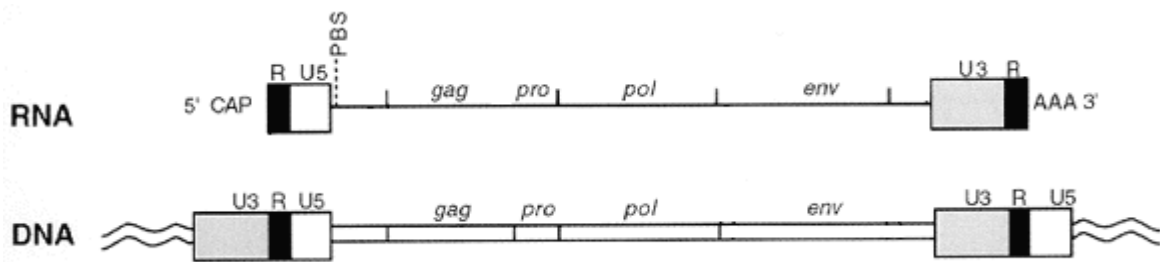


Abb.2: Schematische Darstellung der retroviralen Erbinformation in Form einer mRNA mit 5'Cap, Polyadenylierung am 3'-Ende und Primerbindungsstelle (PBS) für die *reverse Transkriptase* (RT). Darunter befindet sich die in das Genom der Wirtszelle integrierte Form, das Provirus. Während der Integration überschreibt die RT die RNA in eine Doppelstrang-DNA und dupliziert die U3- und U5-Region, so dass funktionsfähige LTRs entstehen. Durch diesen Prozess entsteht ein, verglichen mit der retrovirale RNA, längeres provirales Genom (nach Coffin, 1997).

Nach Integration liegt das Provirus umrahmt von zwei nicht kodierenden Sequenzen, den Long Terminal Repeats (LTRs), vor. Diese analog aufgebauten Regionen, die sich in ihrer Größen innerhalb der Familie der Retroviren stark unterscheiden, bestehen aus drei Elementen. Die U3-Region, die aus dem 3'-Ende der RNA hervorgegangen ist, beinhaltet einen Großteil der Kontrollelemente für Transkription und Replikation. Dazu gehören Bindungsstellen für Transkriptionsfaktoren, Promotor- und Enhancer-Sequenzen. Die Integrität der Enhancer-Sequenzen spielt hierbei eine wesentliche Rolle, da von ihr das Ausmaß der Virusreplikation in der Wirtszelle und damit der Einfluss des Virus auf den Organismus abhängt. Schon eine kleine Veränderung in der Basenfolge dieser Region kann zu einer Veränderung in der Pathogenität eines Virus führen. Die R-Region leitet sich sowohl aus dem 5'- als auch aus dem 3'-Ende der RNA ab, enthält ein Polyadenylierungssignal und spielt eine wichtige Rolle in der reversen Transkription der RNA. Die U5-Region, deren Ursprung im 5'-Ende der RNA liegt, bildet am 3'-Ende das letzte Element eines LTRs. Auch diese Region bedient sich verschiedenster Mechanismen, mit denen sie Einfluss auf die Replikationsaktivität eines Provirus nehmen kann (Rabbi et al., 1997; Kiermer et al., 1998). Im Zuge der reversen Transkription lagert sich eine tRNA an die Primer-Bindungsstelle (PBS) der retroviralen RNA an und stellt ein freies OH-Ende für die Bildung eines DNA-Stranges bereit. Direkt nach dem 5'-LTR folgt stromabwärts das eigentliche retrovirale Genom, das aus Leserahmen für vier Gene besteht. Das *gag*-Gen bildet in 5'-3'-Richtung das erste, gefolgt von den Genen *pro*, *pol* und *env*. Der *gag*-Leserahmen enthält die Information

für das gruppenspezifische Antigen und kodiert für das Gag-Pol-Vorläuferprotein. Aus ihm entstehen nach Transkription und Translation durch proteolytische Spaltung zum einen das Matrix-, das Kapsid- und das Nukleokapsidprotein. Zum anderen bildet der *pol*-Abschnitt die Grundlage für zwei Enzyme. Es kodiert für die *Reverse Transkriptase*, die sowohl für das Umschreiben des Genoms von RNA zu DNA als auch durch ihre *RNaseH*-Aktivität für den Abbau des entstandenen RNA/DNA-Hybridmoleküls verantwortlich ist. Darüber hinaus geht aus dem *pol*-Gen die *Integrase* hervor, die durch ihre Endonuklease- und DNA-Ligaseaktivität das Einschleusen des retroviralen Genoms in die DNA der Wirtszelle ermöglicht. Weiter stromabwärts ist der *env*-Leserahmen lokalisiert, der die Information für die Hüllproteine des Virus enthält. Auch hier entsteht erst ein Vorläuferprotein, das in weitere Proteine, die Transmembran- und die Oberflächenproteine, gespalten wird. Zwischen der *gag*- und der *pol*-Region liegt der Leserahmen für die *Protease*. Je nach Retrovirus wird dieses Protein als Teil des *gag*- oder *pol*-Leserahmens oder in einem separaten Leserahmen exprimiert. Es ist für die proteolytischen Spaltungen während eines retroviralen Lebenszyklus verantwortlich und katalysiert somit die Prozessierung der Gag-, der Pol- und der Env-Vorläuferproteine (Oroszlan und Luftig, 1990). Die Synthese solcher Polyproteine erlaubt es den Retroviren, eine Vielzahl von Genprodukten unter Verwendung einer mRNA zu synthetisieren. Gleichzeitig können dadurch Transportsignale für den gerichteten Transfer in die Regionen von verschiedenen Proteindomänen gemeinsam genutzt werden, in denen die Virusbildung stattfindet.

## 2.2.4 Infektionsverlauf und Replikationszyklus

(*Retroviruses*, Coffin, Hughes, Varmus 1997)

Im ersten Schritt einer retroviralen Infektion heftet sich das Virus mit seinen Oberflächenproteinen (SU) an membranständige Rezeptoren der zukünftigen Wirtszelle. Als Rezeptor fungieren verschiedene Oberflächenmoleküle. Im Fall des HI-Virus ist es der CD4-Rezeptor auf T-Helferzelle und Makrophagen während das Feline Leukemia Virus (FeLV) membranständige Transportermoleküle als Eintrittspforte in die Zelle benutzt (Dalglish et al., 1984; Takeuchi et al., 1992). Dieser Anheftungsprozess unterliegt einer hohen Spezifität und bestimmt das Wirtsspektrum und die Infektiosität eines jeden Retrovirus. Nach der Anheftung vollzieht sich mit Hilfe der viralen Transmembranproteine (TM) der Fusionsprozess, bei dem die Virus- mit der Zellmembran verschmilzt und das Nukleokapsid

in das Zellinnere eindringt. Dort beginnt, katalysiert durch die *Reverse Transkriptase* (RT), die Synthese eines DNA-Einzelstranges (- sssDNA), wobei die RNA als Matrize fungiert und als Resultat ein RNA-DNA-Hybridstrang entsteht. Die RNA-abhängige *DNA-Polymerase* benutzt dazu eine zelleigene tRNA als Primer, die an die Primerbindungsstelle der retroviralen RNA am 5'-Ende bindet. Im weiteren Verlauf dieser Reaktion wird der RNA-Strang durch die *Ribonuklease H* ab- und ein komplementärer DNA-Einzelstrang (+ sssDNA) durch die *RT* aufgebaut. Der so entstandene DNA-Doppelstrang wird in einem nächsten Schritt zusammen mit weiteren Enzymen als Präintegrationskomplex in den Zellkern eingeschleust. Dieser Prozess kann passiv während der Mitose oder aktiv durch Transporter in der Zellmembran des Nukleus ablaufen. Dort angekommen beginnt die eigentliche Integration. Als Voraussetzung dafür entfernt die *Integrase* vorher am 3'-Ende des Doppelstranges die letzten beiden Basen, wodurch dort eine OH-Gruppe freigelegt wird. Diese spaltet am 5'-Ende des DNA-Abschnittes, in den das retrovirale Genom integriert werden soll, die Phosphodiesterbindung. Mit ihrer Endonukleaseaktivität kann sich nun die *Integrase* an die Wirtszell-DNA anlagern und diese so zurechtschneiden, dass der retrovirale DNA-Doppelstrang nahtlos eingefügt werden kann. Bei diesem Prozess entstehen an beiden Enden der Integrationsstelle Lücken, die die *Integrase* mit Hilfe ihrer Ligaseaktivität in einem letzten Schritt wieder auffüllt. Als Resultat dieser Reaktion bleibt das entstandene Provirus im Genom der Wirtszelle zurück, das an beiden Enden von einer für das Retrovirus typischen Sequenz umrahmt wird. In dieser Form persistiert das Provirus und ist, wenn es sich um einfache Retroviren handelt, an die Replikationsaktivität der Zelle gebunden. Bei komplexen Retroviren wie den Lentiviren oder dem Humanen-T-Zell-Leukämie-Virus (HTLV) werden die retroviralen Transkriptionsfaktoren Tat und Tax gebildet, mit deren Hilfe das Provirus einen stärkeren Einfluss auf die zelluläre Replikationsmaschinerie ausüben kann. Zusammen mit Rev und Rex, Proteine, die für den Transport viraler RNA aus dem Nukleus verantwortlich sind, handelt es sich bei diesen Proteinen um Produkte alternativen Spleißens. Unabhängig davon, ob es sich um komplexe oder einfache Retroviren handelt, repliziert die Wirtszelle virale RNA, die, bevor sie ins Zytoplasma gelangt, am 5'-Ende mit einem „Cap“-Zusatz versehen und am 3'-Ende polyadenyliert wird. Nachdem die virale mRNA aus dem Zellkern geschleust wurde, dienen die ungespleißten Moleküle zum einen als Ausgangspunkt für das Gag- und das Gag-Pro-Pol-Vorläuferprotein. Diese Proteine werden myristyliert, an die Innenseite der Plasmamembran transportiert und vermitteln dort die Knospung („budding“) eines neuen Virions aus der Zellmembran. Zum anderen wird RNA, ohne translatiert zu werden, an die Zellmembran transportiert, um dort als virales Genom späterer

Viruspartikel in diese verpackt zu werden. Die gespleißten aus dem *env*-Gen transkribierten mRNA-Moleküle durchlaufen ebenfalls den zellulären Translationsprozess. Die entstandenen Proteine werden dann im rauen endoplasmatischen Retikulum glykosyliert und an die Zelloberfläche transportiert. Dort wandern sie innerhalb der Lipiddoppelschicht an die Stellen, an denen die neuen Viruspartikel ausknospen. Hat sich ein solches Partikel abgeschnürt, beginnt im Inneren der Reifungsprozess. In dessen Verlauf spaltet die virale *Protease* die Vorläuferproteine proteolytisch und überführt sie in ihre aktive Form. Erst jetzt ist das Virion in der Lage neue Zellen zu infizieren.

### 2.2.5 Auswirkungen retroviraler Integration

Da Retroviren auf molekularer Ebene in die genetische Integrität der Wirtszelle eingreifen, besitzen sie ein enormes pathogenes Potential. Es erstreckt sich von Schwächung des Immunsystems über neurologische Erkrankungen bis zur Entstehung von Tumoren. Für das onkogene Potenzial der Retroviren sind zwei verschiedene Mechanismen verantwortlich. Zum einen kann sich die Integration des Provirus zufällig in der Nähe eines zellulären Protoonkogens (*c-onc*) ereignen. Diese Regionen im Genom einer jeden Zelle kodieren für Proteine, die deren Wachstum, Teilung und Differenzierung regulieren. Durch die räumliche Nähe der viralen LTRs mit ihren Replikationsregulatoren kann ein Protoonkogen „an-“ oder „ausgeschaltet“ und damit zu einem Onkogen werden. Die Position des Provirus ist dabei irrelevant, da sich die retroviralen Kontrollelemente sowohl in 5’-, als auch in 3’-Richtung orientieren können. Das Ergebnis ist entweder der Funktionsverlust der Zelle und die damit verbundene Apoptose oder eine verstärkte Replikationsaktivität dieser Region, die sich in vermehrter Zellteilung oder Proteinsynthese äußern kann. Es kommt in der Folge zu keiner direkten malignen Transformation der Zelle, sondern zu einer massiven Replikation viraler Genprodukte über einen längeren Zeitraum. Befallen werden hauptsächlich Zellen der Hämatopoese. Als Beispiel sei das humane T-Zell-lymphotrope Virus 1 (HTLV-1) angeführt, in dessen Infektionsverlauf es zur Entstehung eines Lymphoms kommen kann (Poiesz et al., 2003). Zum anderen kommen Onkogene auch in dem Genom von Retroviren vor und werden dort *v-onc* genannt (Bishop, 1991). Nach Integration dieser meist nicht mehr replikationskompetenten Viren werden ausschließlich die Proteine, für die *v-onc*-Region kodiert, synthetisiert. In Verbindung mit den Replikationsverstärkern in der U3-Region der LTRs haben diese Retroviren die Möglichkeit zur malignen Transformation und können

dadurch in relativ kurzer Zeit Tumoren induzieren. Allerdings ist das Spektrum viraler Infektionen dieser Art begrenzt und die damit verbundene Onkogenese in der Natur eher zu vernachlässigen.

Die durch Retroviren hervorgerufenen neurologischen Erkrankungen zeichnen sich durch zwei verschiedene Verläufe aus. Die Infektion mit dem neurotrope Mäuse-Leukämievirus Cas-Br-E-MLV kann zu einer spongiformen Enzephalopathie führen, die durch den Befall der Neuroglia hervorgerufen wird und ohne Entzündungsreaktion abläuft. Durch Infektion der Gliazellen im Vorderhorn des lumbalen Rückenmarks, des Kleinhirns und im Bereich des Hirnstamms kommt es bei infizierten Mäusen zu einer Paralyse der unteren Extremitäten (Gardner et al., 1973). HTLV-1 ist neben der T-Zell-Leukämie auch für die tropische spastische Paraparese oder HTLV-1 assoziierte Myelopathie (HAM) verantwortlich. Charakteristisch für die HAM ist eine starke Immunreaktion gegen virusinfizierte Zellen des zentralen Nervensystems. Eine wesentliche Rolle spielen dabei  $CD8^+$  T-Zellen, da deren Reaktion mit dem viralen Tax-Protein hauptverantwortlich für die überschießende Entzündungsreaktion zu sein scheint. Als Folge kommt es durch Destruktion des thorakalen Rückenmarkparenchyms zu einer langsam fortschreitenden spastischen Paraplegie der Extremitäten, deren klinischer Verlauf dem der Multiplen Sklerose ähnelt (Jacobsen, 2002). Die weltweit bekannteste retrovirale Infektion, die durch das Humane Immundefizienzvirus (HIV) hervorgerufen wird und in dem erworbenen Immundefizienz-Syndrom (AIDS) mündet, wurde 1984 erstmals beschrieben (Levy et al., 1984; Barre-Sinoussi et al., 1983). Neben dem humanspezifischen Virus ließen sich auch bei anderen Säugetierarten Retroviren nachweisen, die einen ähnlichen Infektionsverlauf zeigten. Der Befall von  $CD4^+$  T-Lymphozyten (T-Helferzellen), die in der Folge stark dezimiert werden, schwächt das Immunsystems des betroffenen Individuums derart, dass er im Verlauf an opportunistischen Infektionen erkrankt, die letztendlich zum Tod führen. Da Lentiviren für diese Form der retroviralen Erkrankung verantwortlich sind, verläuft sie meist über Jahre.

## 2.3 Endogene Retroviren

Gewöhnlich verläuft eine Virusinfektion auf horizontaler Ebene, d.h. die Viren wandern von Wirt zu Wirt und infizieren dort meist ausdifferenzierte Zellen. Integriert sich ein Retrovirus aber in eine Keimbahnzelle und wird deren Erbinformation von einer Generation zur nächsten nach den Mendelschen Gesetzen weitergegeben, wird das virale Genom Teil des

Wirtsgenoms. Solche, in der DNA des Wirtes verankerte Retroviren werden als endogene Retroviren bezeichnet, die in der neuen Generation ubiquitär vorkommen, da sich deren Genom von dem der infizierten Keimbahnzelle ableitet. Darüber hinaus sind endogene Retroviren (ERV) in der Lage, sich durch Transkription, reverse Transkription und erneute Integration intrazellulär zu vermehren. Dieser Retrotransposition genannte Prozess führt zu einer Akkumulation retroviraler Erbinformation im Genom der Wirtszelle. Abhängig vom Zeitpunkt der Integration lassen sich die ERV in „alte“ und „moderne“ Retroviren einteilen. Da sie in allen Mitgliedern einer Klasse nachweisbar sind, fand die Integration der „alten“ ERV vor der Differenzierung der einzelnen Ordnungen statt. Diese Retroviren sind durch Mutationen, Rekombination und Deletionen in ihrem Genom nicht mehr replikationskompetent. „Moderne“ Retroviren finden sich nur in bestimmten Familien oder Gattungen und sind heute vorkommenden exogenen Viren sehr ähnlich, was für eine spätere Integration spricht. Der Nachweis von Genprodukten dieser ERV ist ein Beleg für ihre Replikationsfähigkeit und ein weitestgehend intaktes Genom (Todaro et al., 1974). Ob diese Eigenschaft zur Gefahr für den Träger endogener Retroviren werden kann, ist Gegenstand aktueller Forschung. Bei Mäusen konnte ein Zusammenhang zwischen der Entstehung eines Adenokarzinoms der Brustdrüse und dem Nachweis eines endogenen Retrovirus, das dem exogenen MMTV sehr ähnlich ist, gezeigt werden (Van Nie et al., 1977).

### 2.3.1 Humane endogene Retroviren

In Rahmen der systematischen Sequenzierung des menschlichen Genoms stellte sich heraus, dass dieses zu ungefähr 8% aus Sequenzen besteht, die sich endogenen Retroviren zuordnen lassen (International Human Genome Sequencing Consortium, 2001). Anhand der Unterschiede in der Basenfolge des 5'- und 3'-LTRs eines endogenen Retrovirus und interindividuellen Vergleichen lassen sich Rückschlüsse auf den Zeitpunkt der Integration ziehen (Dangel et al., 1995). Bei Sequenzunterschieden im Bereich von 5-12% und einer Mutationsrate von  $5 \times 10^{-9}$  pro Jahr und DNA-Strang kann man von einem Zeitpunkt ausgehen, der 10 - 25 Millionen Jahre zurückliegt (Steele et al., 1986; Johnson et al., 1999). Das Genom vieler humaner endogener Retroviren (HERV) blieb im Verlauf der Evolution nicht von Mutationen und molekularem Rearrangement verschont. Als Folge sind von den einst intakten Proviren oft lediglich einzelne LTRs oder Genabschnitte nachweisbar. Allerdings konnten auch HERV gefunden werden, deren proviraler Aufbau Bestand hielt. Diese Proviren sind im



Aufbau exogenen Retroviren sehr ähnlich. Diese Tatsache legt die Vermutung nahe, dass diese retroviralen Elemente auch beim Menschen noch aktiv sind und mit dem Organismus des Merkmalsträgers interagieren können. Dass dies nicht zwangsläufig negativ sein muss, zeigt eine Arbeit von Ting et al., in der der Einfluss von HERV-E auf die Amylaseproduktion der Speicheldrüsen beschrieben wird (Ting et al., 1992). Durch Integration stromaufwärts des  $\alpha$ -Amylase-Gens beeinflussen die replikationsregulierenden LTRs des endogenen Retrovirus die Expression des Enzyms. Eine höhere Konzentration an  $\alpha$ -Amylase ermöglicht so dem Merkmalsträger ein schnelleres Erkennen kohlenhydratreicher Nahrung. Blond et al. fanden mit HERV-W ein retrovirales Element, das einen offenen Leserahmen für sein *env*-Gen aufwies und ein Protein exprimierte, das in menschlichem Plazentagewebe gefunden wurde (Blond et al., 1999). Dieses Syncytin genannte Protein spielt eine Rolle in der Bildung von Synzytiotrophoblasten, die als Teil der Endzotten der Plazenta eine wichtige Rolle im maternofetalen Stoffaustausch einnehmen. Ponferrada et al. untermauerten mit einer Arbeit die Hypothese, aktive HERV könnten die Infektion durch exogene Retroviren verhindern (Ponferrada et al., 2003). Gegenstand der Untersuchung war das Env-Glykoprotein einer HERV-W-Sequenz, dessen Einbau in die Zellmembran zu einer deutlich geringeren Infektionsrate durch das Milz-Nekrosevirus (SNV) führte. Kommt das pathogene Potential eines endogenen Retrovirus im infizierten Organismus zum Tragen, übt das einen negativen Evolutionsdruck auf diesen aus. Die Folge ist, dass pathogene Eigenschaften im Laufe der Evolution aus dem Genom entfernt werden. Allerdings gibt es heute Hinweise dafür, dass humane endogene Retroviren in der Entstehung von Multipler Sklerose, Lupus erythematodes, Diabetes mellitus Typ I und anderen Autoimmunerkrankungen eine Rolle spielen könnten (Perron und Seigneurin, 1999). Welchen Einfluss endogene retrovirale Elemente auf die Entstehung maligner Erkrankungen haben, wird intensiv untersucht. So stellen beispielsweise Wang-Johanning et al. in einer ihrer Arbeiten einen Zusammenhang zwischen HERV-E und dem Prostatakarzinom her (Wang-Johanning et al., 2003). Im Rahmen dieser Arbeit wiesen sie eine erhöhte Expression des *env*-Gens in Tumorgewebe, nicht jedoch in gesunden Zellen, nach.

### 2.3.2 Klassifizierung

Die Einteilung der humanen endogenen Retroviren erfolgt in drei Klassen. Als Einteilungskriterium dient eine Sequenzhomologie zu exogenen Retroviren im *pol*-Gen. Der

Klasse I werden all diejenigen HERV zugeordnet, die eine Homologie mit Gammaretroviren aufweisen, wohingegen HERV der Klasse II Basenfolgen mit Alpha-, Beta- und Deltaretroviren gemeinsam haben. Die Klasse III mit HERV-L als bisher einzigem Mitglied weist Homologien zu den *Spumaretrovirinae* auf (Wilkinson et al., 1994). Darüber hinaus kann man innerhalb der Klassen anhand der Bindungsstelle für die tRNA, die die reverse Transkription initiiert, weiter differenzieren. So besitzen die Primerbindungsstellen bei HERV-E, -R, -I, -H und -W jeweils die spezifischen Antikodons für die Aminosäuren Glutamat, Arginin, Isoleucin, Histidin und Tryptophan. An die Primerbindungsstelle der HERV der Klasse II binden nur Lysin-, an die der Klasse III nur Phenylalanin-tRNAs. Aufgrund des Einbuchstabenkodes für Aminosäuren werden diese HERV-K bzw. HERV-L genannt. Diese Klassifizierung ist allerdings wegen teilweise gleichen Primerbindungsstellen zweier verschiedener HERV-Klassen oder bisher nicht bestimmter Bindungsstellen problematisch. Trotzdem wird sie weiterhin genutzt. Innerhalb der Klasse II unterschieden Andersson et al. zwischen zehn weiteren retroviralen Elementen (Andersson et al., 1999). Als Grundlage für diese Differenzierung diente eine Arbeit von Medstrand et al., in der diese Arbeitsgruppe konservierte Regionen der *Reversen Transkriptase* der HERV-K-Proviren mit Sequenzen eines Betaretrovirus verglich (Medstrand et al., 1993). Dabei fanden sich in diesem Bereich deutliche Sequenzhomologien. Es handelte sich bei dem Retrovirus um das Mäusebrusttumorvirus (MMTV), was zu dem Akronym HML (human MMTV like) führte.

Klasse I	Klasse II	Klasse III
<b>HERV-HF</b> HERV-H (RTVL-H, RGH) HERV-F <b>HERV-RW</b> HERV-W HERV-R (ERV9) HERV-P (HuERS-P, HuRRS-P) <b>HERV-ERI</b> HERV-E (4-1, ERVA, NP-2*) HERV-R (ERV3) RRHERV-I <b>HERV-T</b> HERV T (S71, CRTK1, CRTK6) <b>HERV-IP</b> HERV-I (RTVL-I) HERV-IP-T47D (ERV-FTD) <b>ERV-FRD</b>	<b>HERV-K (HML-1)</b> HERV-K (HML-1.1) <b>HERV-K (HML-2)</b> HERV-K10 HERV-K-HTDV <b>HERV-K (HML-3)</b> HERV K (HML3.1) <b>HERV-K (HML-4)</b> HERV-K-T47D <b>HERV-K (HML-5)</b> HERV -K-NMWV2 <b>HERV-K (HML-6)</b> HERV K (HML-6p) <b>HERV-K (HML-7)</b> HERV-K-NMWV7 <b>HERV-K (HML-8)</b> HERV-K-NMWV3 <b>HERV-K (HML-9)</b> HERV-K-NMWV9 <b>HERV-K (HML-10)</b> HERV-KC4	<b>HERV-L</b>

Tabelle 2: Klassifikation humaner endogener Retroviren (nach Nelson et al., 2003)

### 2.3.3 Das humane endogene Retrovirus K

Als Erster beschrieb Ono 1986 eine endogene retrovirale Sequenz im menschlichen Genom, die für alle viralen Gene weitestgehend offene Leserahmen aufwies (Ono, 1986). Ausgangspunkt dieser Entdeckung war die IAP-Sequenz, ein endogenes Retrovirus, das vorwiegend bei Nagetieren zu finden ist. Diese internalisierten A-Partikel leiten sich von Alpharetroviren ab, verfügen über eine erhaltene *pol*-Region und sind in ihrem Aufbau exogenen Retroviren sehr ähnlich. Indem er ein konserviertes *pol*-Gen, das er aus einem Syrischen Goldhamster (*Mesocricetus auratus auratus*) isolierte, als Matrize benutzte, konnte er eine sequenzhomologe Region im menschlichen Genom nachweisen. Des Weiteren lag

diese Sequenz nicht einzeln vor, vielmehr konnten in einem haploiden Genom etwa 50 Kopien dieses HERV-K10 genannten retroviralen Elements gefunden werden. Anhand von Untersuchungen verschiedener Affen- und Primatenarten legten Mariani-Costantini et al. und später Steinhuber et al. einen zeitlichen Rahmen für die erste Integration retroviraler Elemente der HERV-K-Familie fest (Mariani-Costantini et al., 1989; Steinhuber et al., 1995). Grundlage war der Nachweis von HERV-K-Sequenzen bei Altweltaffen (*Catarrhini*), nicht aber bei Neuweltaffen (*Platyrrhini*). Außerdem fanden sich innerhalb der Teilordnung der *Catarrhini* sowohl bei den geschwänzten Altweltaffen (*Cercopithecoidea*) als auch bei den Menschenartigen (*Hominoidea*) derartige Sequenzen. Folglich ereignete sich die Integration nach der Aufteilung in Alt- und Neuweltaffen aber vor der Trennung der Altweltaffen in *Hominoidea* und *Cercopithecoidea*. Die in der Literatur zu diesem Thema genannten Daten schwanken zwischen 25 und 45 Millionen Jahren. Der erste Nachweis für ein weitestgehend intaktes endogenes Retrovirus bei Menschen gelang Mayer et al. (Mayer et al., 1999). Diese Arbeitsgruppe fand auf Chromosom 7 ein retrovirales Element mit offenen Leserahmen für alle proviralen Gene. Für diese HERV-K(HML-2.HOM) genannte Sequenz konnten funktionsfähige Genprodukte nachweisen werden. Bei der proviralen Organisation von HERV-K sind im Aufbau deutliche Parallelen zu exogenen Retroviren zu erkennen (Abb. 3).

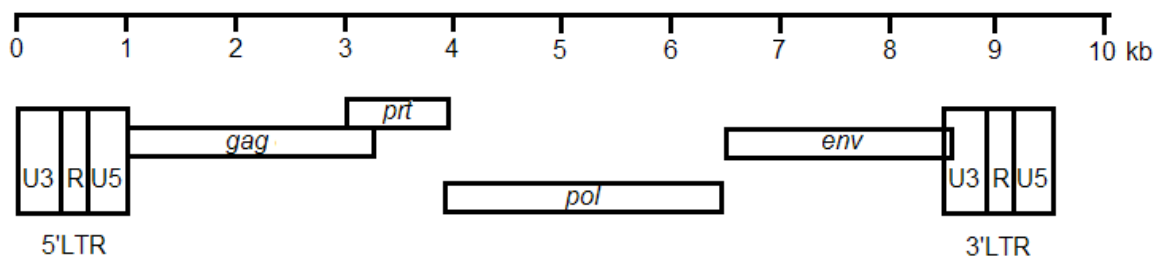


Abb. 3: schematische Darstellung des proviralen Aufbaus von HERV-K (nach Bannert et al., 2004)

Abhängig davon, ob eine 292bp umfassende Sequenz zwischen der *pol*- und der *env*-Region als Teil des proviralen Genoms vorhanden ist, werden alle endogenen Retroviren, die der Familie der HERV-K angehören, in zwei Typen eingeteilt (Löwer et al, 1993). Während Typ-2-Proviren diese Nukleotidfolge aufweisen, kommt es durch deren Deletion bei Typ-1-Proviren zu einer Fusion des *pol*- und des *env*-Gens (Löwer et al., 1993). Innerhalb der deletierten Basenfolge liegt eine Spleißdonorstelle (SD) für eine zweifach gespleißte mRNA, aus der im Verlauf bei Typ-2-Proviren das Rec-Protein translatiert wird (Abb. 4b). Dieses früher cORF genannte Protein spielt analog zu Rev bei HIV und Rex bei HTLV eine wichtige

Rolle im Transport retroviraler mRNA aus dem Zellkern. Es bindet transkribierte Sequenzen, stabilisiert sie und schleust sie aus dem Nukleus, bevor der zelluläre Spleißapparat aktiv werden kann (Magin et al., 1999). Darüber hinaus konnte eine Interaktion mit dem Promyelozytenleukämie Zinkfinger-Protein (PLZF) nachgewiesen werden. Dieses Protein spielt im gesunden Organismus eine wichtige Rolle in der Wachstumsregulation von Zellen und kann bei vorliegendem Defekt zu einer akuten Promyelozyten-Leukämie führen (Shaknovich et al., 1998). Über diese Wechselwirkung scheint Rec transformierende Eigenschaften zu entwickeln, die an der Entstehung von Keimzelltumoren beteiligt sind (Boese et al., 2000). Trotz des Verlustes der Spleißdonorstelle für das Rec-Protein kodiert das Genom der Typ-1-Proviren ebenfalls für ein Protein (Abb.4a), das aus alternativem Spleißen hervorgeht und ausschließlich in entarteten Zellen zu finden ist (Armbrüster et al., 2002). Der SD dieses Np9 genannten Proteins liegt stromaufwärts der Deletionstelle. Ähnlich wie Rec kumuliert es im Zellkern. In einer weiterführenden Arbeit stellten Armbrüster et al. (2004) einen Zusammenhang zwischen diesem Protein und der LNX/Numb/Notch-Signalkaskade her, wobei LNX als Ligand für Np9 zu fungieren scheint. Diese Signalkaskade mit Notch als Rezeptor und LNX und Numb als Liganden spielt in der Embryogenese, der Zelldifferenzierung und –proliferation eine entscheidende Rolle. Die Tatsache, dass ein Fehler in dieser Reaktionskette genau die Entartungen zur Folge haben, die mit Np9 in Verbindung gebracht werden, nämlich Leukämie, Mamma- und Keimzelltumoren, lassen einen Zusammenhang noch glaubhafter erscheinen. Wie genau sich die Interaktion von Np9 mit diesen Proteinen auf die Entstehung von Tumoren auswirkt ist Gegenstand aktueller Forschung.

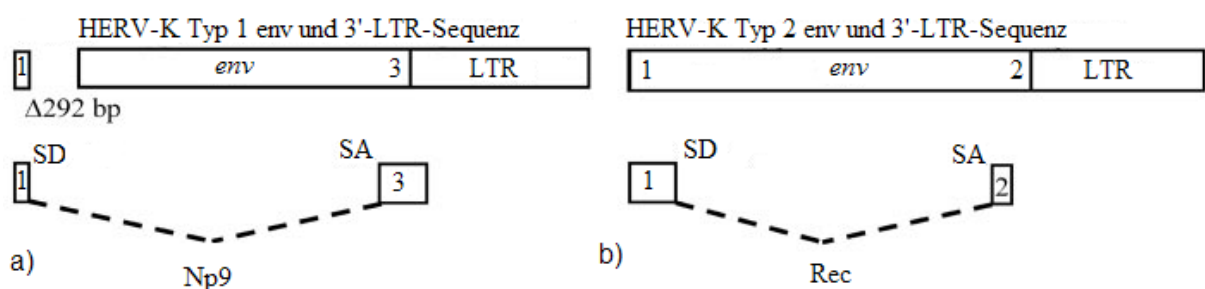


Abb.4: a) Typ 1 Provirus mit Deletion (Δ) und Np9 als Spleißprodukt b) Typ 2 Provirus mit Spleißdonor (SD) und -akzeptor (SA) für Rec (nach Bannert et al., 2004; Armbrüster et al., 2002)

## 2.3.4 Strukturelle Variabilität innerhalb der HERV-K-Familie

Schon die Klassifizierung der humanen endogenen Retroviren lässt erahnen, dass es sich bei HERV-K keinesfalls um eine homogene Gruppe retroviraler Elemente handelt. Vielmehr finden sich heute neben Einzelnukleotidpolymorphismen (SNP) Abweichungen in Form und proviraler Anordnung. Als Passagiere im Genom sind die Proviren den genetischen Rekombinationsmechanismen des Wirtsorganismus ausgesetzt. Allerdings gibt es auch Hinweise auf ein retroviral vermitteltes genetisches Rearrangement während der Evolution des Menschen (Hughes und Coffin, 2001). Die am häufigsten nachzuweisenden Elemente sind die Long terminal Repeats, die durch Rekombination im Genom meist als solo-LTRs vorliegen. Innerhalb dieser Sequenzen lassen sich die meisten Polymorphismen nachweisen (Hughes und Coffin, 2004). Mayer et al. konnten 2005 in einer Haplotyp-Analyse von HERV-K(HML-2) zeigen, dass die Proviren nach der Integration aber auch in Tandem- oder in Einzelform vorliegen, wobei die Tandemformation deutlich häufiger vorgefunden wurde. Diese Ergebnisse bestätigen die Daten einer vorausgegangenen Arbeit, in der die Tandemanordnung für HERV-K(HML-2) als die am weitesten verbreitete angesehen wird (Reus et al., 2001). Zum Nachweis wurden SNPs innerhalb der *gag*-, *pro*- und *pol*-Gene genutzt, die Nonsense-Mutationen zur Folge hatten. Als Mechanismen führt die Arbeitsgruppe Rasterverschiebungen, die Entstehung vorzeitiger Stopkodons und Mutationen im Bereich von Enzymmotiven an. In einer anschließenden Analyse verschiedener ethnischer Gruppen wurden deutliche Unterschiede in der Prävalenz sowohl der SNPs als auch der Tandemformation festgestellt. Bei einer genomweiten Suche nach neuen Mitgliedern der HERV-K(HML-2)-Gruppe stieß eine andere Arbeitsgruppe auf 113 verschiedene retrovirale Elemente, die für den Menschen spezifisch sind (Belshaw et al., 2005). Innerhalb der gefundenen HERV konnten sie acht ausmachen, die nur bei einem Teil der untersuchten Individuen nachweisbar waren. Diese als Insertionspolymorphismus bezeichnete Eigenschaft zusammen mit einer teilweise geringen Allelfrequenz einzelner HERV kann als Hinweis für eine erst kürzlich stattgehabte Integration gedeutet werden. All diese Daten lassen den Schluss zu, dass die für Menschen spezifischen retroviralen Elemente, sei es aus eigenem Antrieb oder durch die Plastizität des Wirtsgenoms, einem ständigen Wandel unterliegen. Barbulescu et al. (1999) diskutieren in ihrer Arbeit den Einfluss genetischer Rekombination auf das Proliferationsverhalten von HERV-K. So könnte etwa durch das Zusammenwirken verschiedener Proviren ein funktionsfähiges Genprodukt oder durch molekulares Rearrangement ein komplettes Provirus entstehen. 2006 gelang es einer französischen

Arbeitsgruppe ein Provirus *in vitro* zu generieren, das sie aus verschiedenen retroviralen Elementen der HERV-K(HML-2)-Gruppe zusammensetzen (Dewannieux et al., 2006). Bemerkenswert darin ist, dass dieses „Phoenix“ genannte Provirus zur Partikelbildung und Reinfektion somatischer Zellen befähigt ist. Damit konnte diese Arbeitsgruppe erstmals zeigen, dass es theoretisch möglich ist, durch Rekombination mehrerer nur teilweise aktiver Proviren, ein voll funktionsfähiges Provirus entstehen zu lassen.

### 2.3.5 Einfluss von HERV-K auf den Wirtsorganismus

Innerhalb der humanen endogenen Retroviren kommen die Mitglieder der HERV-K Familie replikationskompetenten exogenen Retroviren sehr nahe (Löwer et al., 1996). Die damit verbundene Expression auf mRNA- und Proteinebene konnte bereits nachgewiesen werden. So fanden Medstrand et al. retrovirale Ribonukleinsäuren in menschlichem Lungen-, Muskel-, Plazenta- und Nierengewebe (Medstrand et al., 1993). Boller et al. brachten Viruspartikel, die aus einer menschlichen Teratokarzinomzelllinie stammten, mit HERV-K in Verbindung (Boller et al., 1993). Mit Hilfe eines anti-HERV-K-Gag spezifischen Antiserums untersuchten sie durch Immunfluoreszenz und im Elektronenmikroskop diese HTDV (human teratocarcinoma-derived virus) –Partikel. Durch Western Blot konnten zudem HERV-K-Gag-spezifische Proteine in Teratokarzinomzelllinien identifiziert werden. Damit rückten Keimbahntumoren in den Fokus des Interesses. Inhalt zweier Arbeiten von Sauter et al. war die Suche nach Antikörpern gegen HERV-Gag-Proteine bei Patienten mit Keimbahntumoren. Mit dem gelungenen Nachweis konnte eine gerichtete humorale Immunreaktion belegt und darüber hinaus eine deutlich erhöhte Expression dieser Proteine in den Tumorzellen gezeigt werden (Sauter et al., 1995; Sauter et al., 1996). In beiden Arbeiten zeigte sich ein gegenüber gesunden Individuen stark erhöhter Titer der anti-Gag- und der anti-Env-Antikörper. Allerdings kann daraus nicht auf einen ätiologischen Zusammenhang geschlossen werden. Erste Hinweise für einen direkten Einfluss von HERV-K auf die Entstehung von keimbahnassoziierten Tumoren fand die Arbeitsgruppe, als sie die transformierenden Eigenschaften des Rec-Proteins (siehe 2.3.3) bei Mäusen untersuchten (Boese et al., 2000). Im weiteren Verlauf beobachteten Galli et al. bei transgenen Mäusen, die das *rec*-Gen in allen Zellen trugen, nach 12 Monaten Veränderungen im Hodengewebe gefolgt von maligner Entartung nach 19 Monaten (Galli et al., 2005).

Da schon bei der Einteilung der HERV-K-Familie die Homologie zu MMTV-Sequenzen eine wichtige Rolle spielt, liegt die Vermutung nahe, dass es einen Zusammenhang zwischen dem Vorhandensein von HERV-K im menschlichen Genom und der Entstehung von Brustkrebs geben könnte. Wang-Johanning et al. suchten sowohl in gesundem als auch in tumorösem Brustgewebe nach Hinweisen auf eine erhöhte Expression von HERV-K (Wang-Johanning et al., 2001). Im Gegensatz zum gesunden Gewebe fanden sie in fast allen Brustkrebszelllinien und Tumorgeweben *env*-Abschriften. Weitere Analysen ergaben, dass es sich bei den gefundenen Sequenzen um ein HERV-K Typ 1 handelt. 2003 konnte dieselbe Arbeitsgruppe diesen Nachweis auch für HERV-K-Elemente vom Typ 2 erbringen (Wang-Johanning et al., 2003). Dass auch andere gynäkologische Tumoren mit HERV-K und dessen Genexpression in Verbindung gebracht werden können, zeigt eine weitere Arbeit von Wang-Johanning et al. (Wang-Johanning et al., 2007). Neben weiteren endogenen Sequenzen wiesen sie in Ovarialtumorzellen eine erhöhte Expression von Env-Proteinen nach, die eindeutig von einem HERV-K stammten.

### 2.3.6 Die retroviralen Elemente HERV-K113 und HERV-K115

Die retroviralen Elemente HERV-K113 und HERV-K115 nehmen innerhalb der Familie der HERV-K einen Sonderstatus ein. Beide sind Vollängenproviren und heute mit ihren Präintegrationstellen nur bei einem bestimmten Anteil der menschlichen Population zu finden. Turner et al., die diese Proviren erstmals beschrieben, lokalisierten HERV-K113 auf Chromosom 19p13.11 und HERV-K115 auf Chromosom 8p23.1 (Turner et al., 2001). Das Genom von HERV-K115 weist 92bp stromaufwärts des *gag*-Stopcodons eine Deletion von 1bp auf, wodurch der Carboxylterminus des Gag-Vorläuferprotein verändert wird und somit eine Translation der offenen Leserahmen für *pro* und *pol* unwahrscheinlich ist (Abb.5a). Der *env*-Leserahmen dieses Provirus ist unversehrt und kodiert neben einem intakten Hüllprotein auch für das alternativ gespleißte Rec-Protein. Die Besonderheit von HERV-K113 liegt in einem völlig intakten Provirus mit offenen Leserahmen für alle retroviralen Proteine (Abb.5b). Neben der Integrationsstelle, die dieses endogene Retrovirus flankiert, liegt stromabwärts der Insertionsstelle eine duplizierte Region, die zu deren Basenfolge



größtenteils sequenzhomolog ist. Dies konnte Dr. Thomas Stuhr im Rahmen seiner Dissertation anhand von Sequenzanalyse zeigen (Stuhr, 2004).

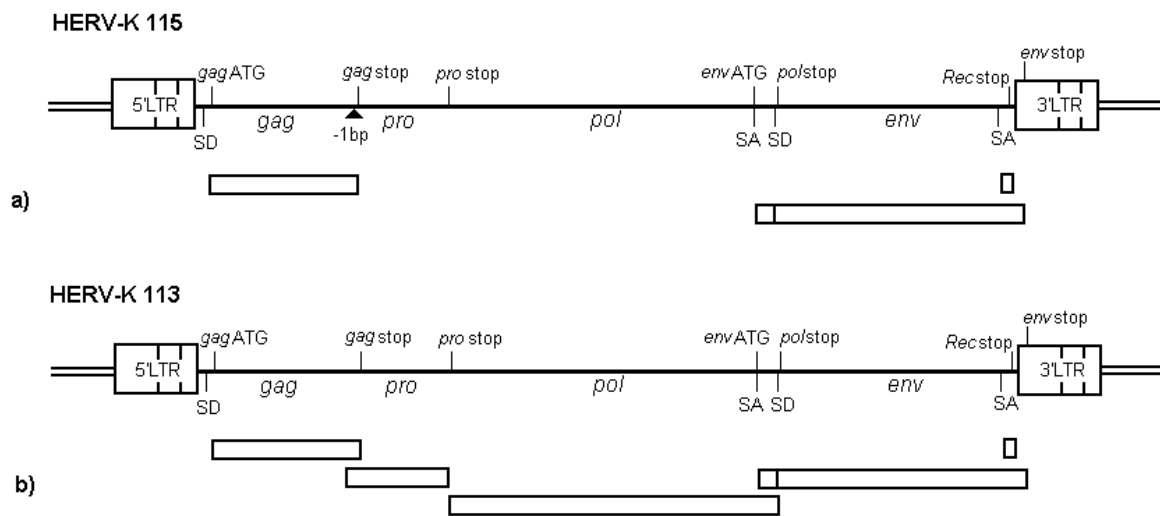


Abb. 5 Proviraler Aufbau von HERV-K113 und HERV-K115 ( nach Turner et al., 2001) a) HERV-K115 Provirus mit zu erwartenden Abschriften von *gag* und *env*, den jeweiligen Start(ATG)- und Stopkodons und Splißdonor(SD)- und Akzeptorstellen(SA) für *rec*; mit  $\Delta$  markierte Deletion am Übergang zwischen *gag* und *pro* b) HERV-K113 mit ORFs für alle proviralen Proteine, deren Start(ATG)- und Stopkodons und den Splißdonor(SD)- und akzeptorstellen(SA) für *Rec*

In einem weiteren Aspekt der Arbeit versuchten Turner et al. den Zeitraum der Integration beider Retroviren festzulegen. Dabei benutzten sie die bereits beschriebene Methode des Sequenzvergleichs der LTRs. Die 14 gefundenen Unterschiede in den beiden LTRs von HERV-K 115, die größtenteils auf Rekombination und Genkonversion zurückzuführen sind, machen eine genaue Altersbestimmung unmöglich. Da dieses Provirus aber humanspezifisch ist, kann sich die Integration erst nach der Entstehung der Gattung Mensch ereignet haben. Zwischen den LTRs von HERV-K 113 waren keinerlei Unterschiede nachweisbar, so dass eine Integration vor weniger als 200000 Jahren als sehr wahrscheinlich gilt.

## 2.4 Ziele der Arbeit

In der Arbeit von Turner et al. von 2001 wurde zum einen die Genfrequenz der retroviralen Elemente HERV-K113 und HERV-K115 in einer Population von 31 Individuen untersucht. Zum anderen konnten sie anhand der jeweiligen Präintegrationsstellen mit Hilfe eines PCR-Schemas klären, ob das HERV-K113-Provirus heterozygot oder homozygot vorlag. Bei der

analysierten Gruppe, die aus Individuen aus Europa, Afrika, Asien und Amerika bestand, fanden sie für HERV-K113 einen Anteil von 29%, für HERV-K115 einen Anteil von 16%. Dies entspricht einer Genfrequenz von 0,15 für bzw. 0,08. In einer weiterführenden Arbeit von Moyes et al. zeigte sich, dass diese stark von der ethnischen Herkunft abhängt und innerhalb verschiedener Krankheitsbilder variiert (Moyes et al., 2005).

Ziel dieser Arbeit soll es im Folgenden sein, eine größere Population auf das Vorhandensein von HERV-K113 und HERV-K115 zu untersuchen und damit ein klareres Bild über die homo- bzw. heterozygote Verteilung dieser Proviren in der menschlichen Population zu erlangen. Dazu soll die zu untersuchende DNA von insgesamt 207 Individuen durch Isolierung über Anionenaustauschersäulen gewonnen und mit Hilfe der Polymerasekettenreaktion und speziell auf die Proviren zugeschnittenen Oligonukleotidprimern untersucht werden. Ein geeignetes PCR-Schema soll erarbeitet und etabliert werden. Aufgrund der Assoziation von HERV-K mit Keimzelltumoren soll überprüft werden, ob bei Patienten mit derartigen Malignomen eine höhere Allelfrequenz der retroviralen Elemente zu beobachten ist, als bei gesunden Individuen. Dies geschieht durch den Vergleich einer Subpopulation von 27 Individuen, die an einem Seminom erkrankt sind, mit einer 80 Individuen umfassenden Kontrollgruppe. Der Nachweis von Homozygotie soll mit Hilfe einer Methode, basierend auf dem Einsatz von Restriktionsendonukleasen, etabliert werden. Im letzten Teil soll durch Sequenzanalyse der *env*- und *gag*-Gene von HERV-K113 verschiedener Individuen die Integrität des proviralen Genoms untersucht und etwaige Polymorphismen analysiert werden. Die dafür benötigte DNA wird als PCR-Fragment in Bakterien vermehrt und über Präparation von Plasmid-DNA im kleinen und großen Maßstab für die Sequenzierung vorbereitet. Darüber hinaus soll diese Untersuchung die Frage klären, ob eventuelle Sequenzunterschiede zufällig verteilt sind oder auf bestimmte Mutationen selektiert werden.

### 3. Material und Methodik

#### 3.1 DNA-Isolierung

##### 3.1.1 Isolierung über Anionenaustauschersäulen

(QIAamp DNA Blood Mini Kit, Fa. Quiagen)

Zur Gewinnung von verwertbarer DNA aus venösem Blut (+ EDTA) wurde auf das Prinzip der Anionenaustauschersäulen zurückgegriffen.

Das venöse Blut stammte zum einen von Mitarbeitern des Institutes, zum anderen wurde es von der Virologischen Diagnostik zur Verfügung gestellt und unter Wahrung der Anonymität weiterverwendet.

Hierbei wurden zu 200µl der Blutprobe erst 20µl Proteinase K, danach 200µl Lysis-Puffer (Puffer AL) gegeben, um die Zellmembranen der Blutzellen zu zerstören, die leukozytäre DNA auszulösen und die zellulären Proteine zu zersetzen. Dieser Vorgang fand bei 56°C im Wasserbad statt, in dem der Ansatz für 15 min inkubiert wurde. Anschließend beendete die Zugabe von 200µl Ethanol (96-100%) die Reaktion. In dem folgenden Schritt band die DNA durch Zentrifugation bei 8000 U/min (Tischzentrifuge) an die positiv geladene Siliziumgelmembran der Säule. Möglich wurde dies durch oberhalb von pH 2 stark negativ geladene Phosphatgruppen am Rückgrat der Nukleinsäuren. Zudem war auch die Salzkonzentration in dem Lysat von Bedeutung für das Binden an die Membran. Sie war so eingestellt, dass sich weniger stark negativ geladene Moleküle wie z.B. RNA oder auch Proteine schlechter an die Matrix anlagern konnten. Da dies aber nicht vollständig verhindert werden konnte, wurde die Säule im Anschluss durch Zugabe von jeweils 500µl zweier Waschpuffer (Puffer AW1 und Puffer AW2) und Zentrifugieren gespült, zuerst bei 8000 U/min für eine Minute, dann für drei Minuten bei 14000 U/min. Aufgrund der aufsteigenden Ionenstärke in den Puffern konnten RNA und Proteine ihre Bindungen nicht länger aufrechterhalten und auf der Membran blieb lediglich die DNA zurück.

Um auch diese von der Säulenmatrix zu trennen konnte man entweder den pH senken oder nochmals die Ionenstärke erhöhen. Auf letzteres wurde zurückgegriffen. In diesem Protokoll wurden 100µl eines Elutionspuffer zugegeben und bevor das Eluat durch Zentrifugieren bei 8000 U/min gewonnen wurde, inkubierte man für einige Minuten bei Raumtemperatur. Dies

sollte laut Hersteller die Ausbeute an DNA, die bei venösem Blut zwischen 4-12 µg liegt, erhöhen. Für sämtliche Zentrifugationschritte bei dieser Methode bediente man sich einer Eppendorf Tischzentrifuge.

### 3.1.2 Isolierung hochmolekularer DNA aus Blut durch Salz-Chloroform-Extraktion

(Mullenbach et al., 1989)

Eine andere Art der DNA-Gewinnung setzte man bei Proben ein, bei denen die Leukozyten durch einen Ficoll-Gradient aufgereinigt waren. Die Zugabe von 9,5ml SE-Puffer resuspendierte die Zellen. Anschließend wurden 0,5 ml Natrium-Dodecylsulfat (SDS) hinzugefügt, so dass eine Endkonzentration von 1% entstand aufgrund derer Enzyme und Proteine, die spätere Reaktionen behindern könnten, lysiert wurden. Bevor der Ansatz über Nacht bei 55°C inkubiert wurde, kamen 100 µl Proteinase K hinzu. Dieses Enzym baut das Chromatingerüst der DNA ab und ermöglicht so einen Zugriff auf die Erbinformation. Die entstandenen Proteinfragmente wurden im Folgenden durch die Zugabe von 4ml 5M NaCl-Lösung (angestrebte Endkonzentration 1,5M), 14 ml Chloroform und anschließendem einstündigen Mischen in einem Überkopfrötator voneinander getrennt. Um die DNA zu erhalten, musste die nukleinsäurereiche wässrige Phase von der proteinreichen öligen Phase durch zehnminütiges Zentrifugieren bei 3000 U/min getrennt werden. Nachdem die DNA in ein neues Gefäß überführt wurde, konnte die Nukleinsäure mittels eines Volumens Isopropanol gefällt werden. Durch den folgenden Zentrifugationsschritt für eine Stunde bei 5500 U/min wurde die DNA pelletiert, anschließend mit 70%igem Ethanol über Nacht gewaschen und erneut bei 5500 U/min für eine Stunde zentrifugiert.

Die gereinigte DNA trocknete bei Raumtemperatur und wurde anschließend in 250 µl TE-Puffer gelöst.

Material:

Zentrifuge: Megafuge 1.0 R, Heraeus

SE-Puffer: 75mM NaCl, 25 mM EDTA pH 8,0

TE-Puffer: 10mM Tris-HCl, 1mM EDTA pH 8,0

10 %iges Natriumdodecylsulfat (SDS)

5M NaCl

Proteinase K 20mg/ml Stammlösung

## 3.2 Photometrische Konzentrationsbestimmung

Für diesen Schritt nutzte man die optischen Eigenschaften von Nukleinsäuren. Durch die Messung der optischen Dichte mit Hilfe der Absorptionsspektroskopie ließ sich sowohl die Konzentration einer definierten DNA-Lösung, als auch deren Reinheitsgrad bestimmen. Nukleinsäuren absorbieren aufgrund der Spektralcharakteristika ihrer Basen Licht im Bereich von 250 nm bis 270 nm. Bei Proteinen liegt dieser Bereich wegen deren hohen Gehaltes an Tyrosin- und Tryptophanresten bei 280 nm. Zur Konzentrationsbestimmung wurde die DNA-Lösung 1:100 mit Aqua bidest. verdünnt und das Wasser als Referenzprobe benutzt. Da es sich bei der zu messenden Nukleinsäure um Doppelstrang-DNS handelte und die Küvette eine Dicke von einem Zentimeter aufwies, errechnete sich nach dem Lambert-Beer'schen Gesetz bei einer Wellenlänge  $\lambda=260\text{nm}$  und einer optischen Dichte von 1,0 eine Konzentration von  $50\mu\text{g/ml}$ . Mit dem so ermittelten Wert wurde eine Lösung mit der Konzentration von  $50\text{ng}/\mu\text{l}$  bei der späteren Polymerase-Kettenreaktion eingesetzt. Um den Reinheitsgrad der gemessenen Lösung zu erhalten, bildete man den Quotienten der Extinktionswerte bei  $\lambda=260\text{nm}$  und  $\lambda=280\text{nm}$ . Lag dieser Wert unter 1,65, war die Probe durch Proteine oder Phenolreste verunreinigt, lag er über 1,85, befand sich entweder RNS oder denaturierte DNS in dem Ansatz. Ein Quotient zwischen 1,65 und 1,85 lies auf eine hinreichend reine DNS-Lösung schließen. Diese Methode wurde ebenfalls zur Konzentrationsbestimmung von Plasmid-DNS nach Maxipräparation benutzt.

### 3.3 Polymerase-Kettenreaktion

( Saiki et.al 1985)

Die im Mittelpunkt dieser Arbeit stehende Methode ermöglicht *in vitro* die Amplifizierung einer spezifischen DNA-Sequenz („template“) unter Zuhilfenahme von DNA-Polymerase, Oligonukleotiden (Primer) und Desoxyribonukleotiden. Prinzipiell besteht die PCR aus drei aufeinander folgenden Schritten, die sich zyklisch wiederholen (Abb. 6). Im ersten Teil der Reaktion denaturiert der DNA-Doppelstrang bei einer Temperatur von ca. 90°C, d.h. es kommt zur Trennung der komplementären Stränge. Durch Abkühlen des Ansatzes auf eine bestimmte Temperatur (ca. 54°C) können sich die Oligonukleotide an die entstandenen Einzelstränge anlagern (annealing). Die Primer genannten Oligonukleotide sind zwischen 20 und 30 Basen groß und binden an zu ihrer Sequenz komplementäre Regionen der Einzelstränge in 5'-3'-Richtung. Im letzten Schritt eines Zyklus kommt es durch Elongation der Einzelstränge in 5'-3'-Richtung zu neuen Doppelsträngen. Als Bausteine dienen Desoxyribonukleotide (dNTPs), die in dem Ansatz in ausreichender Menge enthalten sind. Eine DNA-abhängigen DNA-Polymerase übernimmt den Einbau dieser dNTPs. Da ein Großteil dieser Enzyme allerdings bei Temperaturen oberhalb von 80°C, wie sie im ersten Teil der Reaktion herrschen, inaktiviert werden, greift man hier zu Enzymen, die aus hitzestabilen Bakterien stammen (Innis et al., 1988). In der vorliegenden Arbeit wurde die aus *Thermus aquaticus* isolierte *Taq*-Polymerase benutzt, deren Aktivitätsoptimum bei 72°C liegt und die neben der templateabhängigen noch eine templateunabhängige Aktivität besitzt. Diese besteht darin, dem neu synthetisierten Strang am Ende häufig eine zusätzliche Base anzuhängen, die auf dem Originalstrang nicht vorhanden ist. Meist handelt es sich dabei um ein Adenosin. Nützlich ist diese Base bei der Klonierung von PCR-Fragmenten in einen dafür konstruierten Vektor (z.B. pGEM®), der überhängende Thymidinreste liefert und so den Einbau erleichtert (Marchuck et al., 1991). Denaturierung, Annealing und Elongation wiederholen sich nun zyklisch um die gesuchte DNA-Sequenz zu amplifizieren. Da mit steigender Zyklenzahl allerdings auch die Fehlerrate zunimmt, blieb die Anzahl der Zyklen auf 30 beschränkt.

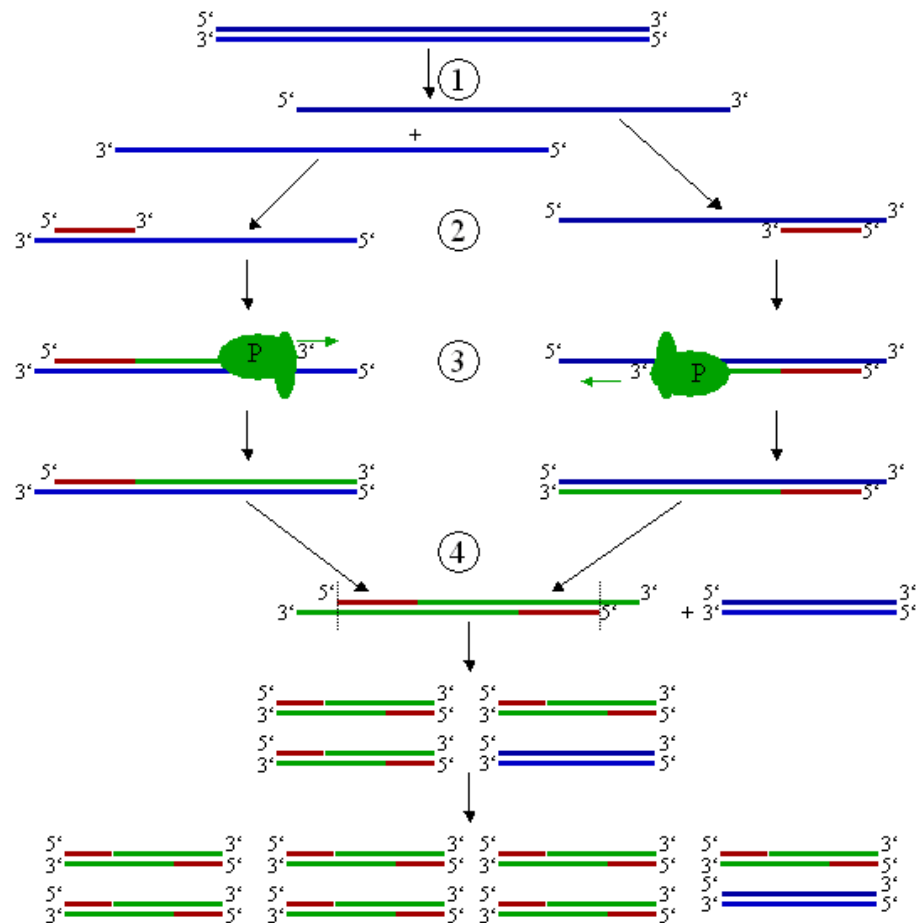


Abb. 6: Phasen der Polymerase-Kettenreaktion

- |                              |      |
|------------------------------|------|
| 1. Denaturierung             | 94°C |
| 2. Annealing der Primer      | 54°C |
| 3. Elongation (P=Polymerase) | 72°C |
| 4. Ende eines Zyklus         |      |

### 3.3.1 Reaktionsbedingungen

Die Zusammensetzung der Reaktionslösung und das Programm, nach dem die Reaktionen abliefen sind in den Tabellen 3 und 4 aufgeführt. Auf die Schemata, mit denen die Polymerasekettenreaktionen durchgeführt wurden um die gesuchten Proviren nachzuweisen, wird im Ergebnisteil dezidiert eingegangen.

Ansatzvolumen	20µl	100µl
Primer A (vorwärts) 25mM	0,3µl	2µl
Primer B (rückwärts) 25mM	0,3µl	2µl
dNTP (0,2 mM)	0,3µl	2µl
Puffer	2µl	10µl
<i>Taq</i> -Polymerase	1 Unit	5 Units
Aqua bidest.	11µl	77µl
DNA (50ng/µl)	6µl	6µl

Tabelle 3: Pipettierschema für den PCR-Ansatz

1. Schritt	3 min	94°C Denaturierung
2. Schritt	50 sec	94°C Denaturierung
3. Schritt	50 sec	58°C Annealing
4. Schritt	3 min	72°C Elongation
5. Schritt	go to 2, 29 mal	
6. Schritt	10min	72°C Elongation
7. Schritt	4 min - ∞	4°C

Tabelle 4: Ablauf des PCR-Programms JM3min mit einer auf die verwendeten Oligonukleotidprimer abgestimmten Annealing-Temperatur ( $T_m$ )

### 3.3.2 Die verwendeten Oligonukleotidprimer

In Zusammenarbeit mit Jens Mayer vom Institut für Humangenetik der Universität des Saarlandes wurden die für diese Arbeit benötigten Oligonukleotide erstellt. Passend zu den durch das Human Genom Project bekannten Sequenzen der untersuchten HERV kam das Primersuchprogramm Primer3 (v.0.4.0) zum Einsatz (Rozen und Skaletsky, 2000). Um die jeweiligen retroviralen Elemente in der Proben-DNA nachzuweisen, sollte je ein Abschnitt des 5' und ein Teil des 3' terminalen Bereichs amplifiziert werden. Außer GAPDH N und GAPDH M wurden alle benutzten Primer auf diese Weise erstellt. Der Tabelle 5 sind die in



dieser Arbeit verwendeten Oligonukleotide mit Annealingtemperatur (T<sub>m</sub>), Sequenz und Accession-Nummer zu entnehmen.

Primer	T <sub>m</sub>	Größe	Sequenz	Accession-Nr.
GAPDH N	56°C	24bp	5'-AGT CCA GTG AGC TTC CCG TTC AGC-3'	B6294G10
GAPDH M	52°C	24bp	5'-TGG TAT CGT GGA AGG ACT CAT GAC-3'	B6294G11
K113 for	58,53°C	27bp	5'-AAA AGC CAT CGA TTG TAC AGT AAT AAA-3'	AF387847
K113 rev	59,66°C	25bp	5'-ACA CAT GCA GAC GTT TAG TTT TGA T-3'	AF387849
env8455 for	59,4°C	20bp	5'-AGA CAG CGA CCA TCG AGA AC-3'	AY037928
LTRgag rev	59,4°C	20bp	5'-TCG AGC GTA CCT TCA CCC TA-3'	AY037928
K113 DuF2	59,05°C	25bp	5'-TAG TCA TCT TGT TAA AAT GAT TTG TT-3'	Chr19:21698880-21698905
K115F1	59,5°C	28bp	5'-ATT TAA AAT TGT CTT CTC CAT CCA CTT T-3'	AF387848
K115r2	59,5°C	26bp	5'-CAG GGT GCA TAC AGC TTA TAA AGA TT-3'	AF387846
ges env for	57,9°C	21bp	5'-ATG AAC CCA TCG GAG ATG CAA-3'	AY037929
ges gag rev	61,0°C	19bp	5'-CAG GCA GGG GGC CAT ATA C-3'	AY037929

Tabelle 5: verwendete Oligonukleotidprimer mit Annealingtemperatur und Accession-Nr. (soweit bekannt)

### 3.3.3 Qualitätskontrolle der isolierten DNA

GAPDH N und GAPDH M sind Oligonukleotide, deren „template“ im Bereich des Gens der Glyzerinaldehyd-3-phosphat-Dehydrogenase liegt. Dieses Enzym spielt eine wichtige Rolle in der Glykolyse und ist in der DNA aller menschlichen Zellen zu finden. Diesen Umstand machte man sich zu Nutze, um die isolierte DNA auf ihre Qualität für die PCR zu testen. Fiel die Reaktion positiv aus, wie in Abbildung 7 zu sehen, konnte man davon ausgehen, dass die DNA nicht degradiert oder übermäßig kontaminiert und für spezifische PCRs geeignet war.

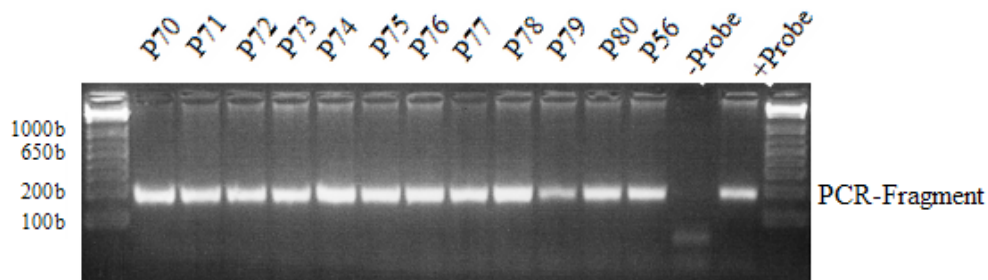


Abb.7: 191bp umfassendes PCR-Produkt nach Reaktion mit den Primern GAPDH N / GAPDH M auf 3%Agarosegel; -Probe: PCR-Ansatz ohne DNA; +Probe: getestete positive DNA

Allen PCR-Ansätzen wurde nach beendeter Reaktion ein Fünftel Volumen Blaumarker zugegeben, um bei der späteren Gelelektrophorese in etwa die Position der DNA-Fragmente bestimmen zu können und etwaige Restaktivitäten der Enzyme zu stoppen.

Dieser Marker besteht aus	Bromphenolblau	0,25% (w/v)
	Glycerin	50% (v/v)
und	EDTA	0,05 M.

### 3.3.4 Nested-PCR

Eine hoch sensitive Form der Polymerasekettenreaktion ist die Nested-PCR. Bei dieser „verschachtelten“ PCR amplifiziert man mit einem Oligonukleotidpaar ein definiertes Fragment und verwendet dieses als template für eine zweite PCR. Die Primer der zweiten Reaktion binden dabei innerhalb des verwendeten Fragmentes. Sinn eines solchen Verfahrens ist es, aus den Produkten der ersten Polymerasekettenreaktion die gesuchten Fragmente spezifisch herauszuselektieren.

## 3.4 Gelelektrophorese

Das Prinzip der Gelelektrophorese besteht darin, die elektrische Ladung von Molekülen zu nutzen. Dabei wandert eine Mischung aus zu trennenden Molekülen unter Einfluss eines elektrischen Feldes durch ein Gel, das sich in einer ionischen Pufferlösung befindet. In Abhängigkeit von Ladung und Größe der Moleküle bewegen sie sich unterschiedlich schnell

durch das als molekulares Netz fungierende Gel. Somit wandern kleine stark negativ geladene Teilchen am schnellsten Richtung Anode, größere, weniger stark geladene Anionen dagegen langsamer. Da es sich bei den DNA-Fragmenten, die in einer Polymerase-Kettenreaktion amplifiziert worden sind, um negativ geladene Moleküle handelt, hat man eine einfache Methode zur Hand, um diese aufzutrennen.

### 3.4.1 Agarose-Gelelektrophorese

Als Hauptbestandteil von Agar, ein aus den Rotalgengattungen *Gelidium* und *Gracillaria* gewonnenes Geliermittel, besteht das Disaccharid Agarose aus D-Galaktose und 3,6-Anhydrogalaktose, die glykosidisch miteinander verbunden sind. Mit Hilfe dieses Zuckers lassen sich Gele unterschiedlicher Konzentration herstellen, abhängig von der aufzutrennenden Fragmentgröße. So weist ein 1%iges Gel z.B. eine Porengröße von etwa 150nm auf und ist damit zur Trennung von DNA-Fragmenten im Größenbereich von 0,5–3kb geeignet. Wählt man eine höhere Konzentration, sinkt die Porengröße und somit steigt das Trennvermögen. Mit einem 3%igen Gel kann man Fragmente bis zu einer Größe von ca. 250 Basen deutlich voneinander unterscheiden. Oberhalb dieser Konzentration stößt diese Methode jedoch an ihre Grenzen. Um ein Agarose-Gel herzustellen benötigt man einen geeigneten ionischen Puffer. Der benutzte TAE-Puffer (Tris-Acetat-EDTA) besaß eine relativ geringe Pufferkapazität, die nur eine Spannung von 0,5-5V/cm Elektrodenabstand erlaubte, da es ansonsten zum Schmelzen des Gels kam. In 400ml diesen Puffers (in einfacher Konzentration) wurden pro Prozentpunkt 4g Agarose durch Kochen in der Mikrowelle gelöst. Bevor es jedoch zum Ausgießen kam, wurden dem Gel 25µl Ethidiumbromid (Stammlösung 100mg/µl) (Abb.8) beigefügt, wodurch eine Konzentration von 6,25µg/ml entstand. Diese Substanz, die zur Infektionsbehandlung und -prophylaxe von Trypanosomen bei Rindern entwickelt wurde, interkaliert zwischen den Basen von DNA und RNA. Durch diese Bindung veränderte sich das Anregungsspektrum und damit die Fluoreszenz der Substanz bei Anregung mit ultraviolettem Licht so stark, dass Nukleinsäurebanden in einem Gel unter der UV-Lampe sichtbar wurden (Le Pecq und Paoletti, 1965).

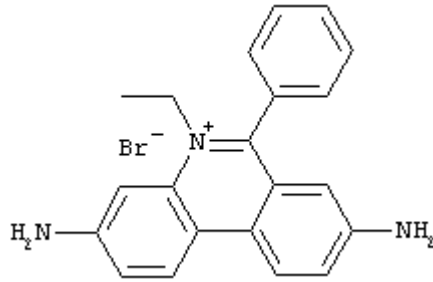


Abb. 8: Strukturformel von Ethidiumbromid

Direkt nach dem Ausgießen in die Form, bevor das Gel anfang zu polymerisieren, erzeugten zuvor installierte Kämme die späteren Pipettiertaschen. Um die DNA-Fragmente einer bestimmten Größe zuordnen zu können, kam der DNA-Molekulargewichtsmarker 1kb + zum Einsatz, der im Gel Banden entstehen ließ, die folgenden Größen (in kb) zugeordnet werden können:

0,1	0,2	0,3	0,4	0,5	0,65	0,85	1	1,65	2	3	4	5	6	7	8	9	10	11	12
-----	-----	-----	-----	-----	------	------	---	------	---	---	---	---	---	---	---	---	----	----	----

Je nach Konzentration des Gels trennte sich entweder das untere oder das obere Spektrum der Banden auf, sodass bei Fragmenten einer Größe von über 1kb ein 1%iges und bei Fragmente unter 0,5kb ein 3%iges Gel benutzt wurde.

### 3.4.2 Polyacrylamid-Gelelektrophorese

Das Agarose-Gel stößt bei kleineren Fragmenten an seine Grenzen. Da es aber in dieser Arbeit darauf ankam DNA-Fragmente im Bereich von 50-100 Basenpaaren differenzieren zu können, wurden Polyacrylamid-Gele in einer Konzentration von 15% benutzt. Für die Herstellung eines solchen Trenngels erfordert

0,8ml Aqua bidest.

4,0ml Trenngelstock (pH 8,8; Tris 1,5 M; SDS 0,4%)

8,0ml Acrylamid (30%)

3,2ml N,N'-Methylenbisacrylamid (2%).

Um aus dem Molekül Acrylamid durch radikalische Polymerisation ein Polymer entstehen zu lassen, wurden dieser Lösung 140µl des Radikalstarters Ammoniumpersulfat und 14µl

Tetramethylethyldiamin, welches als Polymerisierungskatalysator fungierte, hinzugefügt. N,N'-Methylenbisacrylamid quervernetzte die dabei entstandenen linearen Molekülketten und bedingten so die Polymerisation des Gels. Im noch flüssigen Zustand wurde der gesamte Ansatz zwischen zwei Glasplatten gegossen, die mittels dreier Spacern auf Abstand gehalten wurden. Um ein Auslaufen der Lösung zu verhindern, dichtete man den Spalt zwischen den Platten mit 1,5%iger Agarose ab und hielt die Glasscheiben mit Klammern zusammen. Bevor die Polymerisation zu weit fortgeschritten war, wurde am oberen Rand ein Kamm installiert, der, ähnlich wie bei dem Agarosegel, die Taschen bildete, in die später die DNA-Lösung hineinpipettiert werden konnte. Nachdem das Gel nach etwa 30 min vollständig polymerisierte, wurde der Spacer am unteren Rand entfernt und die restliche Konstruktion in eine Gelkammer gespannt. Diese wurde anschließend mit einem Laufpuffer (72g Glycerin, 15g Tris, 25ml SDS 20%, ad 5l Aqua bidest.) befüllt und nach Entfernen des Kammes das Gel beladen. Während der Elektrophorese, die 3-4 Stunden dauerte, sollte auf eine maximale Stromstärke von 20 mA geachtet werden, da ansonsten die Trenneigenschaften des Gels negativ beeinflusst werden. Nach der Elektrophorese musste das Gel angefärbt werden. In dieser Arbeit geschah dies mit einer Ethidiumbromid-Lösung (100 µl auf 200ml TAE-Puffer), in die das Gel ca. 20 min eingelegt wurde. Ein anschließender Waschschriff mit 1%igem TAE-Puffer befreite das Gel von oberflächlicher Ethidiumbromid-Lösung, die auf der UV-Lampe zu Schlieren hätte führen können.

### 3.5 Reisolierung von DNA-Fragmenten aus Agarosegelen

(NucleoSpin® Extract II, Fa. Macherey-Nagel)

Um die auf Agarosegelen aufgetrennten DNA-Fragmente für weitere Schritte nutzbar zu machen mussten diese aus den Gelen herausgelöst werden. Die Firma Macherey-Nagel bietet hierfür ein einfach zu handhabendes Kit an, das, ähnlich wie die Firma Qiagen bei der DNA-Isolierung, auf dem Prinzip der Ionenaustauschersäulen basiert. Im ersten Schritt wurden die Banden, die das gewünschte Fragment enthielten, unter einer Lampe mit langwelligem UV-Licht (366nm) mit einem Skalpell herausgeschnitten und gewogen. Abhängig von der Konzentration des Gels, aus dem die DNA reisoliert werden sollte, und dem Gewicht wurde anschließend bei einprozentigem Gel pro 100mg 200 µl, bei dreiprozentigem die doppelte Menge des NT-Puffers hinzugefügt. In einem Wasserbad (50°C) löste sich anschließend das Gel auf und konnte im Folgenden auf die Siliziumgelmembran in der Säule aufgetragen und

durch Zentrifugieren bei 13000U/min für eine Minute an diese gebunden werden. Der nächste Schritt bestand aus der Zugabe von 600µl NT3-Puffer, eines alkoholischen Waschpuffers, der die Säule durch erneutes Zentrifugieren bei 13000U/min von unerwünschtem Material wie z.B. Primern aus dem PCR-Ansatz befreite. Nachdem der Rest des Waschpuffers durch nochmaliges Zentrifugieren von der Säule entfernt wurde, gab man zur Elution den leicht alkalischen NT-Puffer (ca. 30 µl) hinzu und inkubierte die Säule für eine Minute bei Raumtemperatur. In einem letzten Zentrifugationsschritt bei 13000U/min löste sich die gereinigte und reisolierte DNA von der Säule und wurde für weitere Schritte bei 4°C im Kühlschrank gelagert. Auch hier kam zum Zentrifugieren eine Eppendorf Tischzentrifuge zum Einsatz.



Abb. 9: Prinzip der Reisolierung von DNS-Fragmenten aus Agarosegelen (nach Fa. Macherey-Nagel)

### 3.6 Spaltung von DNA durch Restriktionsendonukleasen

Hinter dem Begriff Restriktionsendonukleasen verbergen sich Enzyme, die in der Lage sind an DNA zu binden und diese zu spalten. Es wird zwischen drei Kategorien unterschieden. Typ-I-Enzyme bestehen aus drei Untereinheiten, die die DNA-Sequenz erkennen, sie methylieren oder schneiden. Das Erkennen der Sequenz findet zwar spezifisch statt, die Stelle, an der das Enzym schneidet, ist aber willkürlich. Typ-III-Enzyme bestehen ebenfalls aus mehreren Untereinheiten, die Schnittstelle liegt aber auch hier 20-30 Nukleotide von der Bindungsstelle entfernt. Somit wird deutlich, dass diese beiden Untergruppen für einen molekularbiologischen Einsatz nur bedingt nützlich sind. Der große Vorteil der Typ-II-Restriktionsendonukleasen ist deren Fähigkeit, spezifisch an DNA zu binden und an selber Stelle zu schneiden. Es handelt sich in diesem Fall um zwei unabhängige Proteine, Restriktionsenzym und Methylase, die beide die gleiche Zielsequenz erkennen. Das eine schneidet, während das andere hemimethylierte DNA vollständig methyliert. Dieser Schritt ist notwendig, da ansonsten das Restriktionsenzym die DNA vollständig verdauen würde. Die

nach Spaltung der Phosphodiesterbindung entstehenden Enden besitzen entweder überhängende Basen (sticky ends) oder sind glatt (blunt ends). In der heutigen Molekularbiologie verwendete Nukleasen, wurden aus verschiedenen Bakterien isoliert und jedem Enzym kann eine bestimmte Bindungssequenz und Schnittstelle zugeordnet werden. Sind diese Eigenschaften und die Art, wie die Nukleasen schneiden, bei verschiedenen Enzymen gleich, spricht man von Isoschizomeren. Liegt der Unterschied lediglich in der Art der entstehenden Schnittstelle handelt es sich um Neoschizomere. Entsprechend des Gattungs- und Artnamens des Bakteriums, aus dem eine Nuklease isoliert wurde, setzt sich die Bezeichnung der Enzyme zusammen. So stammt z.B. das Enzym *AflIII* von *Anabaena flos-aquae*. Die Ordnungszahl gibt die Anzahl der bereits isolierten Enzyme aus einem Bakterium an. Die Einheit, mit der die Konzentration der Restriktionsendonukleasen bezeichnet wird, ist eine Unit. Darunter versteht man die Menge an Enzym, die benötigt wird, um innerhalb von 60min ein  $\mu\text{g}$  der angegebenen DNA bei optimalen Rahmenbedingungen in einem 50 $\mu\text{l}$ -Ansatz vollständig zu verdauen. Die in dieser Arbeit verwendeten Restriktionsendonukleasen und die dazugehörigen Puffer wurden von der Firma New England Biolabs bezogen und sind Tabelle 6 zu entnehmen. Je nach Menge der eingesetzten DNA wurden für einen Verdau zwischen 2 und 3 Units eingesetzt und der Ansatz für mindestens zwei Stunden bei 37°C inkubiert. Jedem Enzym kann ein bestimmter Puffer zugeordnet werden, der für das optimale chemische Milieu sorgt. Manche Nukleasen benötigten außerdem noch acetyliertes Rinderserumalbumin (BSA), das deren Aktivität nochmals erhöhen sollte.

Enzym	Bindungssequenz	Puffer	BSA	Konzentration
<i>Afl</i>	5'...CTTAAG...3' 3'...GAATT <sub>▲</sub> C...5'	NE-Puffer 2/4	100 µg/ml	10000 U/ml
<i>NotI</i>	5'...GC <sub>▼</sub> GGCCGC...3' 3'...CGCCGG <sub>▲</sub> CG...5'	NE-Puffer 3	100 µg/ml	20000 U/ml
<i>SacI</i>	5'...GAGCT <sub>▼</sub> C...3' 3'...C <sub>▲</sub> TCGAG...5'	NE-Puffer 1/4	100 µg/ml	20000 U/ml
<i>SacII</i>	5'...CCGC <sub>▼</sub> GG...3' 3'...GG <sub>▲</sub> CGCC...5'	NE-Puffer 4	-	5000 U/ml
<i>SphI</i>	5'...GCATG <sub>▼</sub> C...3' 3'...C <sub>▲</sub> GTACG...5'	NE-Puffer 1/2/4	-	5000 U/ml
<i>SspI</i>	5'...AAT <sub>▼</sub> ATT...3' 3'...TTA <sub>▲</sub> TAA...5'	NE-Puffer SspI	-	5000 U/ml

Tabelle 6: benutzte Enzyme mit Schnittstelle, passendem Puffer und eingesetzter Konzentration  
(Fa. New England Biolabs)

Zusammensetzung der verwendeten Puffer:

- 1xNE-Puffer 1: 10mM Bis-Tris-Propan-HCl, 10mM MgCl<sub>2</sub>, 1mM Dithiothreitol, pH 7,0
- 1xNE-Puffer 2: 10mM Tris-HCl, 50mM NaCl, 10mM MgCl<sub>2</sub>, 1mM Dithiothreitol, pH 7,9
- 1xNE-Puffer 3: 50mM Tris-HCl, 100mM NaCl, 10mM MgCl<sub>2</sub>, 1mM Dithiothreitol, pH 7,9
- 1xNE-Puffer 4: 20mM Trisacetat, 50mM K-Acetate, 10mM Mg-Acetate, 1mM Dithiothreitol, pH 7,9
- 1xNE-Puffer SspI: 100mM Tris-HCl, 50mM NaCl, 10mM MgCl<sub>2</sub>, 0.025% TritonX-100, pH 7,5



### 3.7 Vermehrung von PCR-Fragmenten durch Bakterien

Der Sinn dieses Verfahrens ist die Gewinnung großer Mengen identischer DNA-Klone unter Zuhilfenahme kompetenter Bakterien. Kompetent bedeutet, dass die Bakterien in die Lage versetzt worden sind, Plasmid-DNA aus der Umgebung aufzunehmen. Zum Einsatz kam in dieser Arbeit pMOS, ein Stamm von *Escherichia coli*-Bakterien. Diese Methode besteht aus drei Arbeitsschritten, der Klonierung, der Transformation und dem Screening.

#### 3.7.1 Klonierung

(pGEM®-T Vector, Fa. Promega)

Hierunter versteht man das Einbringen einer DNA-Sequenz, auch Insert genannt, in einen bakteriellen Vektor. Als ein solches Transportvehikel für Nukleinsäuren werden Plasmide, ringförmige doppelsträngige DNA-Moleküle, die unabhängig von deren Genom in Bakterien vorkommen, genutzt. Verwendet wurde hier der pGEM®-T Vektor der Firma Promega (Abb.10), der mit einer multiple cloning site, einer 126bp umfassenden Region, versehen ist, die Schnittstellen für 15 Restriktionsendonukleasen bietet.

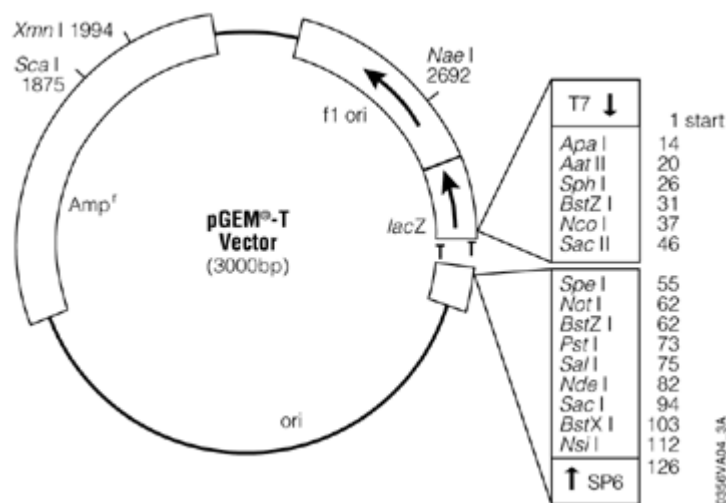


Abb.10: pGEM®-T Vektor

Dadurch kann das Insert später, ohne dabei selbst geschnitten zu werden, aus dem Vektor entfernt werden. Ein Resistenzgen und das LacZ-Gen sind für das Screening nach positiven Klonen ebenfalls vorhanden. Um im Folgenden die DNA-Sequenz in den Vektor klonieren zu

können, wurde ein Ligationsansatz bestehend aus T4-DNA-Ligasepuffer 2x, T4-DNA-Ligase, dem Vektor und dem gereinigten PCR-Produkt hergestellt und dieser bei 4°C über Nacht inkubiert.

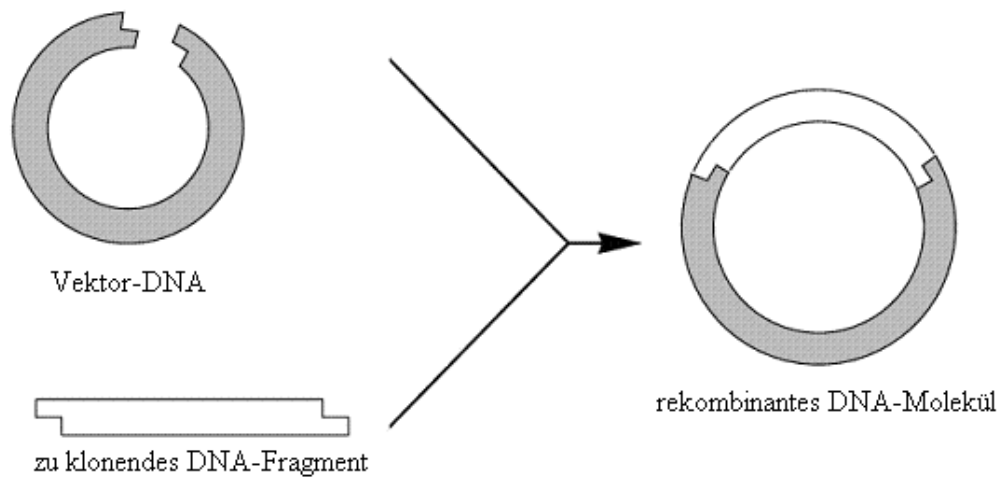


Abb.11: schematische Darstellung der Klonierung

Den Einbau der DNA-Sequenz in den Vektor katalysiert das Enzym *Ligase*. Der benötigte Puffer schafft dafür die optimalen Reaktionsbedingungen. Die genauen Mengen sind der Tabelle 7 zu entnehmen.

Ansatzvolumen	20 µl
T4-DNA-Ligasepuffer	10 µl
T4-DNA-Ligase	6 Units
pGEM-T Vektor	2 µl
PCR-Produkt	6 µl

Tabelle 7: Pipettierschema für Ligationsansatz

### 3.7.2 Transformation

(Cohen und Chang, 1974)

Bei diesem Vorgang wird fremde DNA in Bakterien eingeschleust. In diesem Fall sollten pMos-Bakterien mit dem pGEM®-T Vektor, der durch Klonierung mit dem gewünschten Insert versehen wurde, transformiert werden.

Um dies zu erreichen mussten in einem ersten Schritt die Bakterien, die bei  $-70^{\circ}\text{C}$  lagern, eine Stunde auf Eis auftauen. Im Folgenden konnte man nun  $50\mu\text{l}$  der Bakteriensuspension auf den Ligationsansatz geben. Dieser blieb danach für 20min auf Eis stehen und wurde für einen Temperaturschock für zwei Minuten bei  $42^{\circ}\text{C}$  inkubiert. Wieder auf Eis gestellt wurde dem Ansatz  $200\mu\text{l}$  LB-Medium zugegeben und die so entstandene Vorkultur für eine Stunde bei  $37^{\circ}\text{C}$  unter Schütteln inkubiert. Die transformierten Bakterien wurden anschließend auf einem Nährboden ausplattiert und über Nacht bei  $37^{\circ}\text{C}$  bebrütet (Abb.12). Dem Nährboden setzte man zusätzlich ein Antibiotikum, in diesem Fall Ampicillin, bei. Da auf dem Vektor ein Resistenzgen gegen diesen Stoff vorhanden war, konnte man so bereits während der Inkubation selektiv die transformierten Bakterien anzüchten.

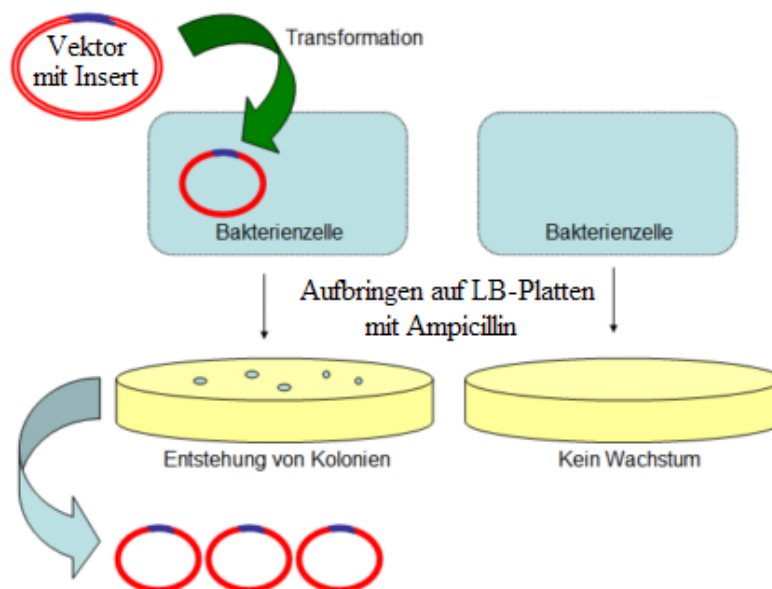


Abb.12: Durch das Einbringen des mit dem Insert ligierten Vektors in Bakterienzellen werden diese transformiert und anschließend auf einem Nährboden, der mit Ampicillin versetzt wurde, angezüchtet

Bevor dies geschah, wurde auf die Petrischale mit dem Nährboden  $20\mu\text{l}$  IPTG und  $35\mu\text{l}$  X-Gal ausplattiert, die für das spätere Screening nach positiven Klonen von Bedeutung waren.

Die Substanzen ließ man für mindestens 30 min bei Raumtemperatur einwirken. Welche Rolle diese Lösungen spielten, ist dem nächsten Abschnitt zu entnehmen.

Material:

LB-Medium:	10g/l Trypton, 5g/l Hefeextrakt, 10g/l NaCl
LB-Nährboden:	LB-Medium mit 1,5% Agar versetzt, Ampicillin 1:1000 (100mg/ml)
IPTG:	100mM Stammlösung, in Aqua bidest. gelöst, bei -20°C
X-Gal:	Stammlösung (c=50mg/ml) in Dimethylformamid, bei 4°C

### 3.7.3 Screening positiver Klone

(Maas, 1999)

Zur Selektion der transformierten Klone diente das Blau-Weiß-Screening, bei dem diejenigen Bakterienklone, die den Vektor mit dem Insert aufgenommen hatten eine weiße, die anderen eine blaue Farbe annahmen. Grundlage für dieses Verfahren ist das lacZ'-Gen auf dem pGEM®-T Vektor, welches für das N-terminale  $\alpha$ -Fragment der  $\beta$ -Galaktosidase kodiert. Dieses Fragment allein besitzt allerdings keine ausreichende Enzymaktivität um das Monosaccharid abzubauen. Nur in Verbindung mit dem C-terminalen  $\omega$ -Fragment, welches im Genom kompetenter Bakterien vorhanden ist, kann Galaktose verstoffwechselt werden (Ullmann et al., 1965). Im Rahmen dieses als  $\alpha$ -Komplementation bezeichneten Vorgangs wird der Stoff X-Gal (5-Brom-4-chlor-3-indoxyl- $\beta$ -D-galactopyranosid) von der intakten  $\beta$ -Galaktosidase abgebaut und es entsteht ein blauer Farbstoff, das 5-Brom-4-chlor-Indoxyl, welches mit Sauerstoff zum tiefblauen Farbstoff 5,5'-Dibrom-4,4'-dichlor-Indigo oxidiert. Die Aufgabe des Isopropyl- $\beta$ -D-thiogalactopyranosids (IPTG) besteht in der Aktivierung des Laktose-Operons, der Funktionseinheit bestehend aus Promotor, Operator und Strukturgen, die für die Transkription der  $\beta$ -Galaktosidase verantwortlich ist.



Abb.13:a) 5-Brom-4-chlor-3-indoxyl-β-D-galactopyranosid b) Isopropyl-β-D-thiogalactopyranosid

Wie der Abb.10 zu entnehmen ist, liegt die multiple cloning site genau in dem lacZ'-Gen, was bei erfolgreicher Klonierung und Transformation dazu führt, dass die Proteinbiosynthese der β-Galaktosidase nicht zustande kommen kann. Somit stellen sich diejenigen Bakterienkulturen, die das gewünschte DNA-Fragment in sich tragen, weiß dar (Abb.14). Je nach Größe des Inserts kann es allerdings dazu kommen, dass auch bei erfolgreicher Klonierung eine Blaufärbung einzelner positiver Kolonien zu beobachten ist. Gründe hierfür sind zum einen in der Länge der Sequenz zu sehen. Je kleiner die Sequenz ist, desto wahrscheinlicher tritt dieses Phänomen ein. Zum anderen könnte ein Fusionsgen dafür verantwortlich gemacht werden. Um diesem Umstand zu berücksichtigen, wurden neben den weißen ebenfalls einige blaue Kolonien abgetragen und erneut inkubiert.

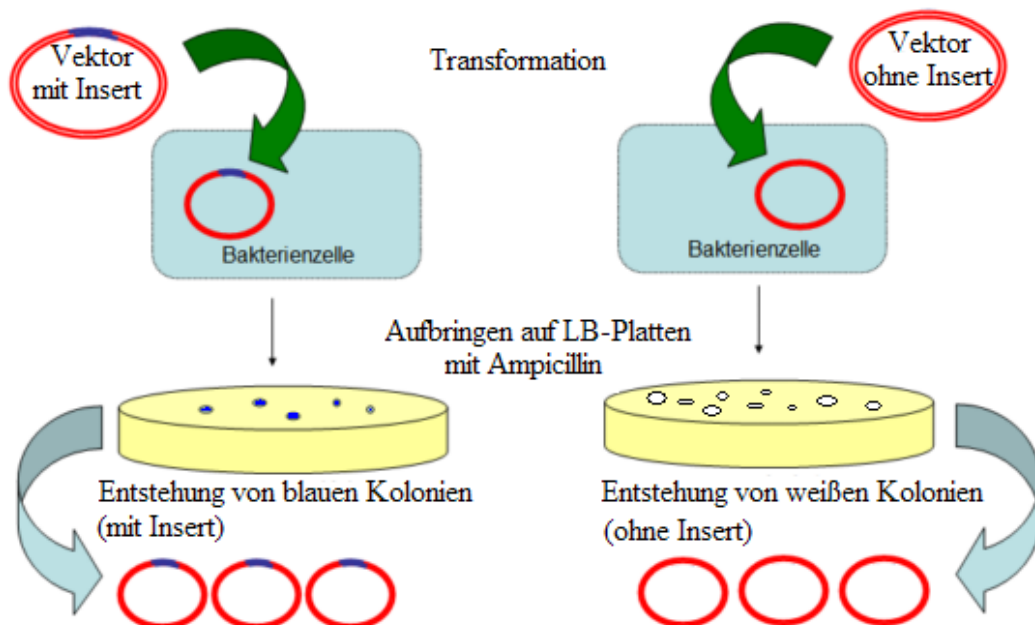


Abb.14: Schematische Darstellung des Blue-White-Screenings von Bakterien, die mit pGEM (mit und ohne Vektor) transformiert wurden

### 3.8 Präparation von Plasmid-DNA in kleinem Maßstab

(Holmes et al., 1981)

Mit dieser Methode wird die Plasmid-DNA aus den Bakterien herausgelöst und für weitere Arbeitsschritte gereinigt. Damit eine hinreichend große Menge an Bakterien zur Verfügung steht, wurden die einzelnen Kolonien in 1,5ml LB-amp-Medium aufgenommen und über Nacht bei 37°C unter Schütteln inkubiert. Diese Kulturen wurden am folgenden Tag für 5min bei 5000U/min zentrifugiert und das Pellet mit 200µl Stetpuffer+Lysozym resuspendiert. Damit sich die Plasmide aus den Bakterien lösen konnten, kam die Suspension im nächsten Schritt für 30sec in kochendes Wasser und danach sofort auf Eis. Nach erneutem 15 minütigen Zentrifugieren bei 4°C und 13000U/min konnte ein erneut gebildetes Pellet bestehend aus Bakterienmaterial mit einem Zahnstocher entfernt werden. Die Plasmid-DNA, die man mit 200µl Isopropanol bei 4°C und 13000U/min für 15min fällte, wurde dann mit 70%igem Ethanol gewaschen. Nachdem das Ethanol dekantiert wurde, musste das Pellet bei Raumtemperatur mindestens 10 min trocknen bevor es in 30µl TE-Puffer+RNase A wieder suspendiert werden konnte.

Material:

Stetpuffer: EDTA (50 mmol/l); Triton X 100 (5% v/v); Sucrose (250mmol/l);  
TrisHCl, pH 8 (50mmol/l); ad 100ml Aqua bidest.

Lysozym: Stammlösung 50mg/ml, in Aqua bidest. gelöst

LB-amp-Medium: 10g/l Trypton; 5g/l Hefeextrakt; 10g/l NaCl; Ampicillin 1:1000  
(100mg/ml)

TE-Puffer+RNase: 10mM Tris-HCl; 1mM EDTA, pH 8,0; RNase A (10mg/ml)

### 3.9 Kontrollverdau

Da man nicht davon ausgehen konnte, dass sich in jedem gereinigten Plasmid das gewünschte Insert befindet, sollte ein weiterer Kontrollschritt diese Frage klären. Der benutzte pGEM-Vektor mit seiner multiple cloning site mit Schnittstellen für diverse

Restriktionsendonukleasen ermöglichte den Einsatz der Enzyme *SacI* und *SacII*, deren Schnittstellen der Abb.15 zu entnehmen sind.

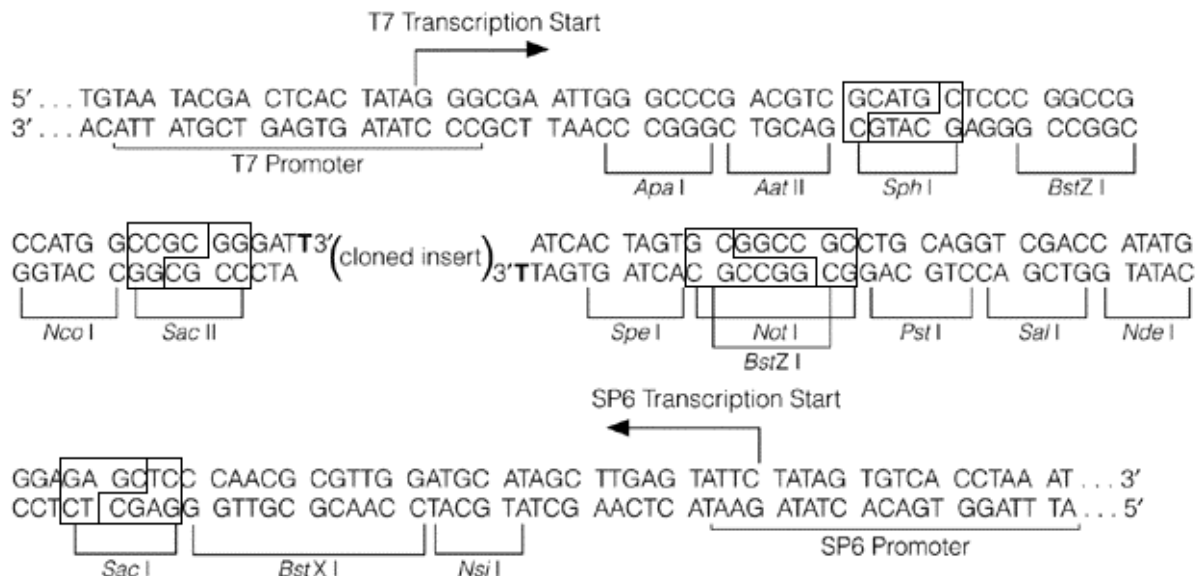


Abb.15: multiple cloning site des pGEM-Vektors mit den dazugehörigen Enzymschnittstellen

*SacI* stammt aus einem *E.coli*-, *SacII* aus *Streptomyces lividans*-Stamm, die jeweils das entsprechende Gen von *Streptomyces achromogenes* in sich tragen. Beide wurden in einer Konzentration von 20U/μl benutzt. Die Zusammensetzung des Ansatzes für den Verdau zeigt Tabelle 8. Der ausschlaggebende Grund für die Wahl dieser beiden Enzyme ist die Tatsache, dass beide nicht in dem eingesetzten Insert schneiden.

Ansatzvolumen	20μl	Ansatzvolumen	20μl
<i>SacI</i> (20U/μl)	1 μl	<i>NotI</i> (20U/μl)	1 μl
<i>SacII</i> (5U/μl)	1 μl	<i>SphI</i> (5U/μl)	1 μl
Puffer	2 μl	Puffer	2 μl
BSA (100μg/ml)	0,5 μl	BSA (100μg/ml)	0,5 μl
DNA (50ng/μl)	5 μl	DNA (50ng/μl)	5 μl
Aqua bidest.	11 μl	Aqua bidest.	11 μl

Tab.8: Pipettierschema für den Orientierungsverdau mit *SacI*, *SacII* und *NotI* und *SphI* und den jeweils benötigten Agenzien

Abhängig von der Größe des Inserts ließen sich nun die Längen der auf einem Agarosegel zu erwartenden Banden berechnen. Da die Schnittstelle von *SacI* 43bp und die von *SacII* 6bp von der Integrationsstelle entfernt liegt, konnte man bei einer Insertlänge von 261bp mit einer Bande bei 310bp rechnen. Eine zweite Bande bei 2599bp repräsentierte den pGEM-Vektor, aus dem das Insert herausgeschnitten wurde. Abb.16 zeigt diesen Kontroll exemplarisch. War kein Insert vorhanden, war das Produkt des Verdaus 48bp lang und kaum nachzuweisen. Auf diese Weise konnte man positive Klone bestimmen, die sich für ein erneutes Anzüchten eigneten.

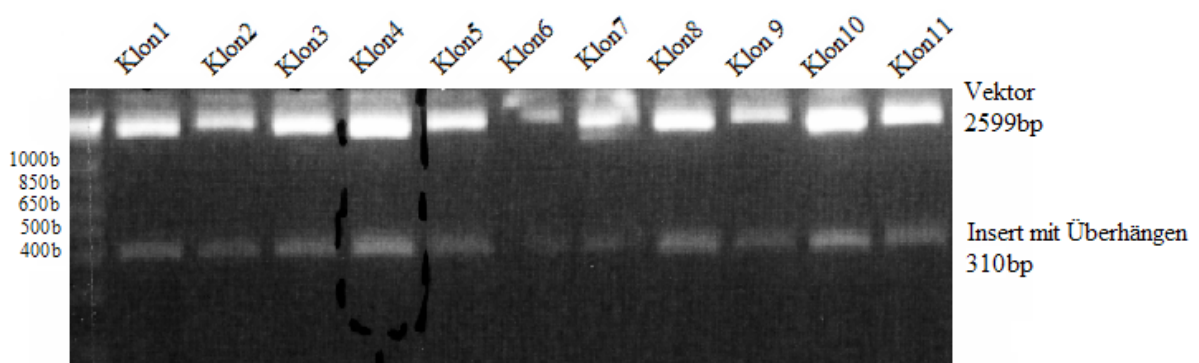


Abb.16: Beispiel eines Kontrollverdaus. Die Abb. zeigt die enzymatische Spaltung von pGEM mit der Integrationsstelle als inseriertes PCR-Produkt durch *SacI* und *SacII* auf 3% Agarosegel.

### 3.10 Präparation von Plasmid-DNA in großem Maßstab

(Maxipräparation mittels NucleoBond® AX 100, Fa. Macherey-Nagel)

Ziel dieser Methode ist die Gewinnung großer Mengen gereinigter Plasmid-DNA, die ein gewünschtes Insert enthält. In dieser Arbeit waren es ungefähr 100µg, da dies die Grenze der Bindungskapazität der verwendeten Anionenaustauschersäulen darstellte.

Im ersten Schritt setzte man eine Kultur pMOS-Bakterien an, transformierte und züchtet sie erneut an. Um die kultivierten Bakterien weiter verarbeiten zu können wurde der Ansatz zuerst für 15 min bei 4000 U/min und 4°C zentrifugiert (Megafuge 1.0 R), der entstandene Überstand dekantiert und das Zellpellet lysiert. Diese geschah durch Zugabe von jeweils 4 ml der Puffer S1 (+RNase A), S2 und S3. Zusammensetzung und Konzentration der einzelnen Puffer ist Tabelle 9 zu entnehmen. Anschließend filtrierte man das Lysat mit einem Papierfilter. Bevor der Ansatz auf die Anionenaustauschersäule gegeben werden konnte, musste die Säule mit Hilfe des N2-Puffers an dessen chemisches Milieu angeglichen werden.



S1: 50 mM Tris / HCl, 10 mM EDTA, 100 µg RNase A / ml, pH 8.0  
S2: 200 mM NaOH, 1% SDS  
S3: 2.80 M KAc, pH 5.1  
N2: 100 mM Tris, 15% ethanol and 900 mM KCl adjusted with H<sub>3</sub>PO<sub>4</sub> to pH 6.3, 0.15% Triton X-100  
N3: 100 mM Tris, 15% ethanol and 1150 mM KCl adjusted with H<sub>3</sub>PO<sub>4</sub> to pH 6.3  
N5: 100 mM Tris, 15% ethanol and 1000 mM KCl\* adjusted with H<sub>3</sub>PO<sub>4</sub> to pH 8.5

Tabelle 9: Verwendete Puffer und deren Zusammensetzung

Nachdem die in der gereinigten Lösung enthaltene Plasmid-DNA an die Säule gebunden hatte, folgte ein Waschschriff mit 10ml N3-Puffer. Anschließend wurde die gebundene und gewaschene DNA mit 5ml N5-Puffer von der Säule gelöst und mit Isopropanol bei 6000 U/min und 4°C in der Megafuge 1.0 R (Fa. Heraeus) für 30 min gefällt. In einem letzten Waschschriff wurden 2 ml 70%iges Äthanol zu dem erhaltenen Pellet gegeben und mit 6000 U/min für 10 min zentrifugiert. Der Überstand musste daraufhin vorsichtig entfernt und das Pellet bei Raumtemperatur 10-15 min trocknen.

Zur Rekonstitution der Plasmid-DNA löste man den getrockneten Rückstand in 100µl Aqua bidest. auf. Zur anschließenden photometrischen Konzentrationsbestimmung wurde die Lösung im Verhältnis 1:100 mit Aqua bidest. verdünnt. Ein letzter Kontrollschritt sollte zeigen, dass sich das PCR-Fragment auch nach der Wiederaufbereitung noch im Vektor befand. Die Abb.17 zeigt diesen Schritt am Beispiel des *env*-Gens, das hier als Insert diente. Durch den Verdau mit der Restriktionsendonuklease *SacI* stellte sich bei erfolgreicher Präparation auf dem Agarosegel zum einen der das Insert enthaltene linearisierte Vektor dar. Dessen Größe lag bei einer Länge des ligierten Fragments von etwa 3100bp bei ungefähr 6100bp. Zum anderen schnitten die Enzyme *NotI* und *SphI* das Insert aus dem Vektor. Da das *env*-Gen innerhalb seiner Sequenz eine zusätzliche Schnittstelle für *SphI* aufweist, wird dieses Insert von *SphI* und *NotI* aus dem Vektor und von *SphI* nochmals in dessen Verlauf geschnitten. Somit beweist der Verdau mit diesen Endonukleasen das Vorhandensein des gewünschten Inserts.

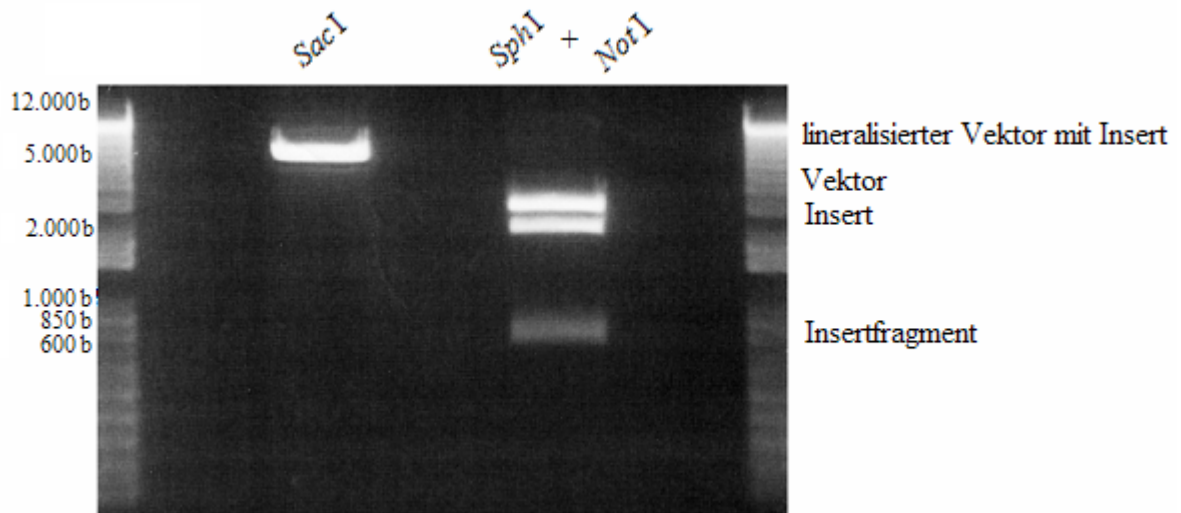


Abb.17: Verdau der Plasmid-DNA durch die Restriktionsendonukleasen *SacI*, *NotI* und *SphI*. Als Insert diente hier das *env*-Gen von HERV-K113 mit einer Länge von ca. 3100bp. Damit ergibt sich für den linearisierten Vektor eine Größe von etwa 6100bp. Der Vektor ohne das Insert hat eine Länge von 2990bp, das Insert 2321bp und das durch die *SphI*-Schnittstelle entstandene Fragment 789bp.

### 3.11 Sequenzierung von DNA-Fragmenten

(Fa. GATC-Biotech, Konstanz)

Um die Nukleotidfolge des zu untersuchenden DNA-Moleküls zu bestimmen, wurde die durch die Maxipräparation gewonnene Plasmid-DNA auf eine Konzentration von 0,1 µg/µl gebracht und in einem Volumen von 30 µl zum Sequenzieren an die Firma GATC-Biotech geschickt.

## 4. Ergebnisse

### 4.1 Zusammensetzung der untersuchten Population

Der mengenmäßig größte Anteil der untersuchten Gruppe stammte aus der hauseigenen Virologischen Diagnostik und belief sich auf 122 Proben. Weitere 36 steuerten Mitarbeiter des Instituts bei und mit den 49 Proben (Bam1–Bam49) aus der Pathologie in Bamberg ergab sich eine Gesamtzahl von 207, die auf das Vorhandensein von HERV-K113 und HERV-K115 untersucht werden sollten. Die Anonymität der untersuchten Proben blieb gewahrt und alle in dieser Arbeit durchgeführten Analysen beschränkten sich auf die DNA. Bei den betrachteten Personen handelte es sich ausnahmslos um Menschen kaukasischer Herkunft. Die geschlechtsspezifische Verteilung lag bei 78 weiblichen und 80 männlichen Individuen. Die Proben aus Bamberg stammten von männlichen Patienten, die an einem keimbahnassoziierten Tumor litten. Bei 27 von diesen handelte es sich um ein Seminom. Unter den verbliebenen 158 (P1–P158) untersuchten Personen konnten anamnestisch folgende Auffälligkeiten festgestellt werden:

- ein Plasmozytom (P94)
- ein Mamma-Ca (P28)
- ein Bronchial-Ca (P15)
- ein Hirntumor (P30)
- zwei Non-Hodgkin-Lymphome (P85; P139)
- ein Tumor nephrolog. Genese (P147)

### 4.2 Die verwendeten PCR-Schemata

Für die in dieser Arbeit durchgeführten Polymerase-Kettenreaktionen kamen verschiedene Primerpaare zum Einsatz, deren Einsatz im Folgenden genauer beschrieben wird.

## 4.2.1 Nachweis von HERV-K113

### Der 5' terminale Bereich

Die beiden Primerpaare K113 for / LTRgag rev und env8455 for / K113 rev dienten zum Nachweis von HERV-K113. Der Vorwärtsprimer des ersten Paares band am 5'-Ende in der flankierende Sequenz von HERV-K113, der Rückwärtsprimer am 5'-Ende des *gag*-Bereichs, so dass das 5'-LTR komplett im „template“ enthalten war. Das hierbei zu erwartende Produkt hatte eine Größe von 1191 Basenpaaren (Abb. 18a).

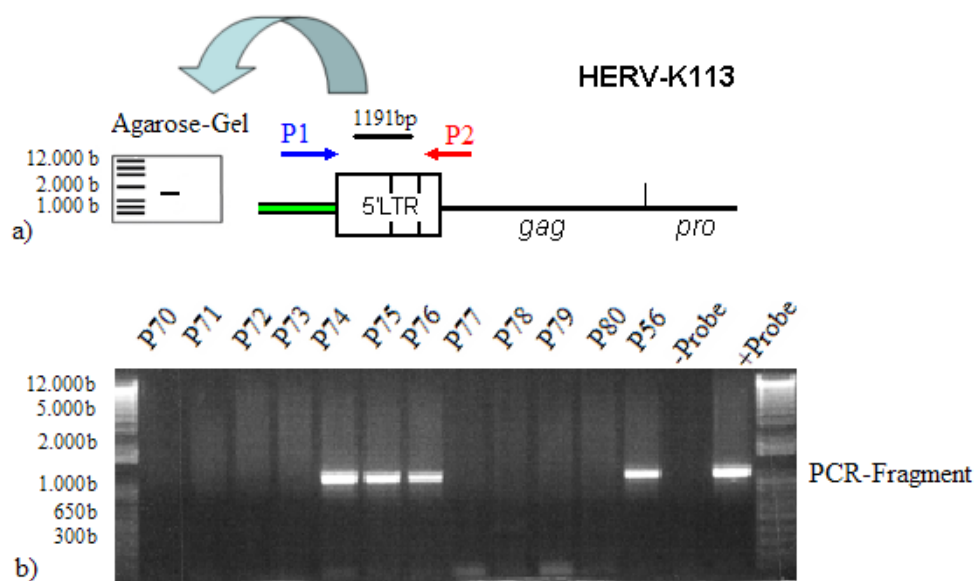


Abb.18: Fragmente nach der PCR mit den Primern K113 for (P1) / LTRgag rev (P2); a) schematische Darstellung der Primerbindungsstellen, der Präintegrationsstelle und dem zu erwartenden Produkt bei positiven Merkmalsträgern; b) 1% Agarosegel mit den Ergebnissen nach der Polymerasekettenreaktion für 12 Proben exemplarisch gezeigt; -Probe: PCR-Ansatz ohne DNA; +Probe: getestete positive DNA

Die so durchgeführte PCR erbrachte innerhalb der untersuchten Population mit 207 Individuen 28 positive Ergebnisse. In der Abb.18b ist das Ergebnis für 12 Proben exemplarisch dargestellt. Die Proben P74, P75, P76 und P56 sind positiv und tragen somit den 5'terminalen Bereich von HERV-K113.

### Der 3' terminale Bereich

Mit dem zweiten Oligonukleotidpaar (env8455 for / K113 rev), von dem der Vorwärtsprimer am 3'-Ende des *env*-Bereichs band und der Rückwärtsprimer sich an die flankierende Sequenz am 3'-Ende von HERV-K113 anlagerte, wurde das 3'-LTR und ein Teil des

stromabwärts liegenden Bereiches erfasst. Hierbei entstand ein PCR-Produkt mit einer Länge von 1157 Basenpaaren (Abb. 19a).

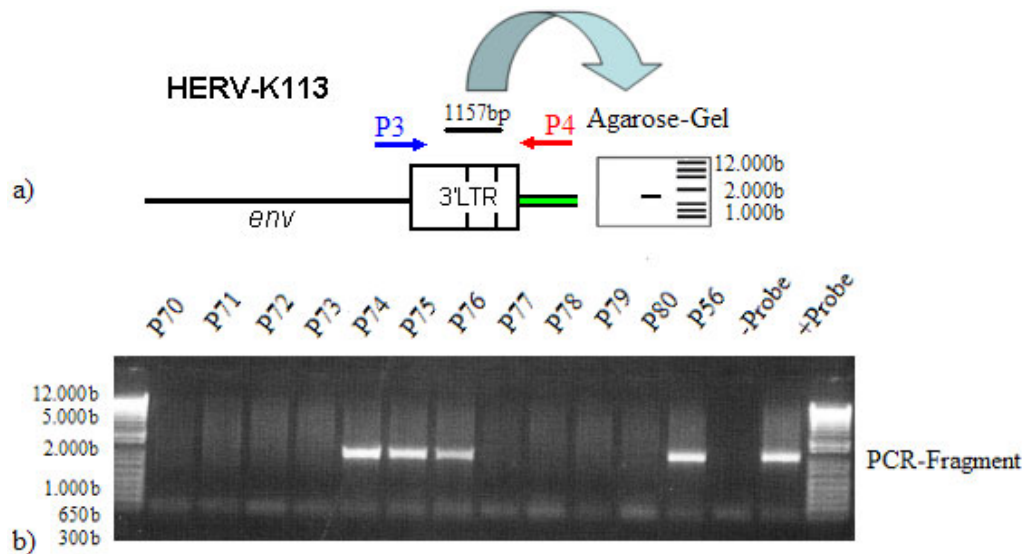


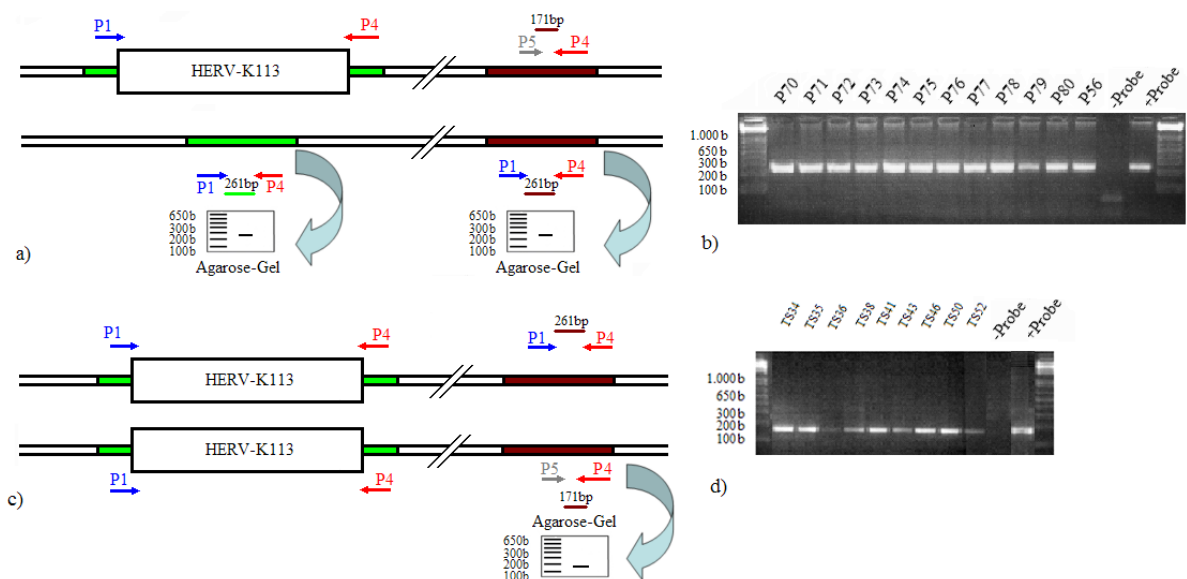
Abb.19: Fragmente nach der PCR mit den Primern *env*8455 for (P3) / K113 rev (P4); a) schematische Darstellung der Primerbindungsstellen, der Präintegrationsstelle und dem zu erwartenden Produkt bei positiven Merkmalsträgern; b) 1%Agarosegel mit den Ergebnissen nach der Polymerasekettenreaktion für 12 Proben exemplarisch gezeigt; -Probe: PCR-Ansatz ohne DNA; +Probe: getestete positive DNA

Die Abb.19b stellt beispielhaft das Ergebnis einer solchen Polymerasekettenreaktion dar. Der Einsatz dieses Primerpaares erbrachte innerhalb der 207 untersuchten Proben analog zu K113for / LTRgag rev ebenfalls 28 positive Ergebnisse. Das gezeigte PCR-Ergebnis weist damit die 3'terminale Region von HERV-K113 für die untersuchten Proben nach, deren 5'terminaler Bereich in Abb.18 dargestellt werden konnte. Somit wurde das integrierte Provirus an beiden Enden nachgewiesen.

#### 4.2.2 Analyse der Präintegrationsstelle und der duplizierten Region

In einem weiteren Ansatz wurden die beiden Primer K113 for und K113 rev kombiniert, um die Präintegrationsstelle und die duplizierte Region näher charakterisieren zu können. Letztere ist, wie in Kapitel 2.3.6 erwähnt, eine der Präintegrationsstelle nahezu identische Region, die jedoch nie mit einem HERV-K113-Provirus besetzt vorgefunden wurde. Das hieraus hervorgehende Produkt hat, unabhängig davon, ob es sich um die Präintegrationstelle oder die

duplizierte Region handelt, eine Länge von 261 Basenpaaren. Aufgetragen auf ein Agarosegel kann man folglich nicht differenzieren, um welche Sequenz es sich handelt (Abb.20a/c). Zur weiteren Charakterisierung wurden die PCR-Produkte von positiven Spendern in den Vektor pGEM ligiert, in Bakterien hochgezogen und nach Aufarbeitung der DNA zur Sequenzierung vorbereitet. Die Sequenanalysen von 10 positiven Klonen ergaben jedoch lediglich Sequenzen der Präintegrationsstelle. Um die duplizierte Region nachweisen zu können wurde eine nested-PCR angewendet, die spezifisch nur die duplizierte Region erfassen sollte. Dazu wurden die Primer K113DuF2 und K113 rev kombiniert, einem Aliquot (1µl) aus dem PCR-Ansatz mit den Primern K113 for / K113 rev zugesetzt und eine PCR durchgeführt. Der Primer K113DuF2 bindet spezifisch innerhalb der duplizierten Region. Aufgrund der Annealingtemperatur und der Stringenz sollte nur die duplizierte Region und nicht die Präintegrationsstelle amplifiziert werden. Das Ergebnis ist ein Fragment mit 171bp. Die PCR zeigte bei allen Analysen ein positives Ergebnis (Abb.20).



Ergebnis der PCR mit dem Primerpaar K113 for K113 rev für 12 Proben gezeigt (Abb.20b). Zum anderen erkennt man eine 171bp große Bande als Ergebnis der PCR mit dem Oligonukleotidpaar K113 DuF2 und K113 rev (Abb.20d). Die Fragmente wurden wiederum kloniert und sequenziert. Mit Hilfe weiterer PCR und Sequenzanalysen konnte die gesamte duplizierte Region nachgewiesen werden. Sie ist im Anhang aufgeführt (siehe Kap. 7.1).

### 4.2.3 Nachweis von HERV-K115

Zum Nachweis des humanen endogenen Retrovirus K115 wurden die Primer LTRgag rev und env8455 for, die zum Nachweis von HERV-K113 dienten, übernommen und mit zwei neuen Oligonukleotiden, die in den flankierenden Bereichen von HERV-K115 binden, kombiniert. Möglich ist dies, da die Sequenzen der *gag*- und der *env*-Leserahmen bei beiden Retroviren soweit homolog sind, dass die Primer sich an die Regionen beider Retroviren anlagern können.

#### Der 5' terminale Bereich

Zum einen kam erneut LTRgag rev zum Einsatz, hier allerdings in Verbindung mit K115 r2. Dieser Primer band in der flankierenden Sequenz des 5'-Endes von HERV-K115 (Abb.21a). Das zu erwartende Fragment bei Vorliegen des Provirus hatte eine Größe von 1100bp.

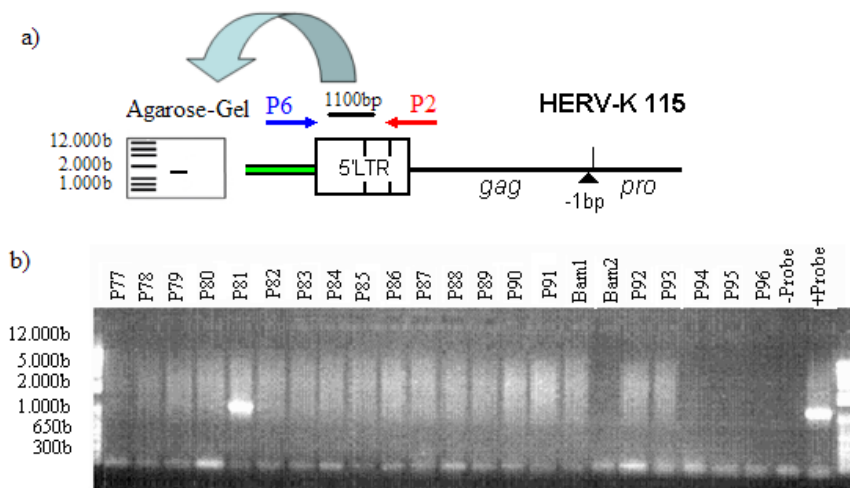


Abb.21: PCR mit den Primern K115r2 (P6) / LTRgag rev (P2); a) schematische Darstellung der Polymerasekettenreaktion mit Bindungsstellen der Primer und der flankierenden Sequenz; b) Ergebnis einer PCR mit Darstellung der Fragmente auf einen 1%-Agarosegel; -Probe: PCR-Ansatz ohne DNA; +Probe: getestete positive DNA

Innerhalb der analysierten Population mit 207 Proben konnten mit Hilfe dieser Reaktion neun positive Merkmalsträger gefunden werden, die in Abb. 21b exemplarisch für 22 Proben gezeigt wurde. Die Probe P81 repräsentiert dabei einen positiven Merkmalsträger.

#### Der 3' terminale Bereich

Analog zu HERV-K113 rev wurde env8455 for in diesem Ansatz mit K115f2 kombiniert. Dieses Oligonukleotid lagerte sich in der flankierenden Sequenz am 3'-Ende des humanen endogenen Retrovirus K115 an und ergab im positiven Fall ein Fragment mit 1250bp (Abb.22).

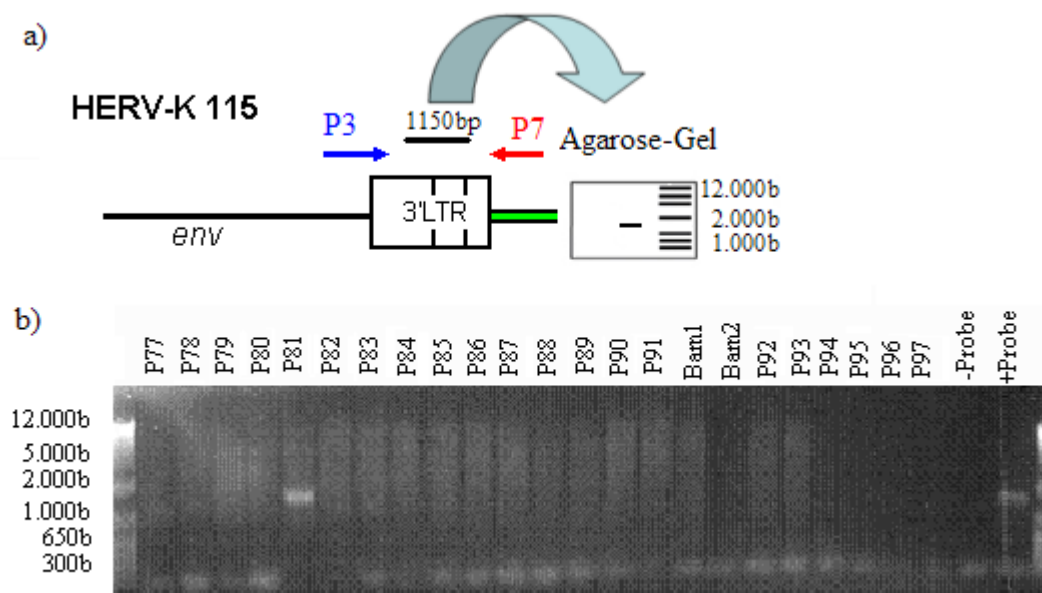


Abb.22: PCR mit den Primern env8455 for (P3) / K115f2 (P7); a) schematische Darstellung der Polymerasekettenreaktion mit Bindungsstellen der Primer und der flankierenden Sequenz; b) Ergebnis einer PCR mit Darstellung der Fragmente auf einen 1%-Agarosegel; -Probe: PCR-Ansatz ohne DNA; +Probe: getestete positive DNA

Die neun bereits gefundenen Merkmalsträger konnten mit dieser Polymerasekettenreaktion bestätigt werden. Analog zu Abb. 21b findet sich auf dem auf Abb.22b dargestellten 1%Agarosegel für die Probe P81 eine Bande.

Mit der Auswahl der in dieser Arbeit benutzten Primer konnte sowohl der N- als auch der C-terminale Bereich der untersuchten humanen endogenen Retroviren nachgewiesen und untersucht werden. Damit konnte bei allen positiven Merkmalsträgern das Vorliegen eines Volllängenprovirus bestätigt werden.



### 4.3 Verteilung der retroviralen Elemente HERV-K113 und HERV-K115

Mittels der etablierten Screening-Methode konnten in der beschriebenen Gruppe 28 positive Merkmalsträger für HERV-K113 und 9 für HERV-K115 gefunden werden. Bezug nehmend auf die Gesamtzahl ergibt sich damit ein prozentualer Anteil von 13,5% für HERV-K113 und 4,4% für HERV-K115 (Tabelle 10). Die Genfrequenz, also die relative Häufigkeit der Kopien eines Allels in einer Population unabhängig von deren Verteilung im Genpool, ist ebenfalls angegeben. Sie berechnet sich aus der Zahl der Kopien eines bestimmten Allels dividiert durch die Gesamtzahl der Kopien aller Allele, die in der Population vorhanden sind. In der vorliegenden Arbeit bedeutet das beispielsweise für die Genfrequenz von HERV-K113 bei 28 Merkmalsträgern und 414 vorhandenen Allelen eine Genfrequenz von 0,07. Allerdings gilt dies nur für heterozygote Merkmalsträger, da diese pro diploidem Chromosomensatz nur ein Merkmal tragen. Bei homozygoten Individuen verdoppelt sich dementsprechend die Zahl der Allele. Bei der Berechnung der hier aufgeführten Genfrequenzen für HERV-K113 sind die Analysen der homozygoten Individuen (siehe Kap. 4.4) einbezogen worden.

	Anteil abs.	Anteil rel.	Genfrequenz
HERV K113	28/207	13,5%	0,07
HERV K115	9/207	4,4%	0,02

Tabelle 10: nachgewiesene Häufigkeit bzw. Genfrequenz

Somit liegt die in dieser Arbeit nachgewiesene Häufigkeit unter den 30% bzw. 15%, die bisher in den Vorarbeiten vermutet wurden. Bestätigt hat sich allerdings die verhältnismäßige Verteilung beider retroviraler Elemente untereinander. Auch die bisherigen Untersuchungen zeigten, dass HERV-K113 in etwa zwei bis dreimal so häufig vorkommt wie HERV-K115.

### 4.3.1 Vergleich der Häufigkeit innerhalb der untersuchten Gruppe

Da in der Literatur ein Zusammenhang zwischen humanen endogenen Retroviren und der Entstehung von keimbahnassoziierten Tumoren – und hier vor allem Seminomen – beschrieben worden ist, wurden in der vorliegenden Arbeit die nachgewiesenen Häufigkeiten in den jeweiligen Gruppen miteinander verglichen. Bei den untersuchten Gruppen handelte es sich zum einen um die Seminomgruppe mit einer Fallzahl von 27 und um die Kontrollgruppe mit einer Fallzahl von 80. Da sich die Betrachtung nur auf Seminome und nicht auf andere keimbahnassoziierte Erkrankungen beschränken sollte, wurden eben diese aus der ersten Gruppe herausgenommen und in die Kontrollgruppe nur Proben männlicher Individuen eingeschlossen. In der Kontrollgruppe waren 15 Personen positiv für HERV-K113 und 6 Personen positiv für HERV-K115, womit sich eine Häufigkeit von 18,8% bzw. von 7,5% ergab. Die Seminomgruppe wies drei Merkmalsträger für HERV-K113 und zwei für HERV-K115 auf, was einer Häufigkeit von 11,1% bzw. 7,4% entspricht. Auch hier ermittelte man wieder die Genfrequenzen, die neben den anderen Daten der Tabelle 11 zu entnehmen sind.

	HERV-K113			HERV-K115		
	positiv	Häufigkeit	Genfrequenz	positiv	Häufigkeit	Genfrequenz
Seminomgruppe n=27	3	11,1%	0,06	2	7,4%	0,04
Kontrollgruppe n=80	15	18,8%	0,09	6	7,5%	0,03

Tabelle 11: Häufigkeitsverteilung von HERV-K113 und HERV-K115 in den untersuchten Gruppen

Bei Betrachtung der jeweiligen Häufigkeit lässt sich ein Trend erkennen, der den direkten Einfluss beider humanen endogenen Retroviren auf die Onkogenese des Seminoms als eher unwahrscheinlich erscheinen lässt.

#### 4.4 Der Nachweis von Homozygotie bei dem HERV-K113-Provirus

Zur Untersuchung auf Homo- bzw. Heterozygotie wurden zu den in der vorliegenden Arbeit positiven 28 Proben weiter 18 einbezogen, die aus der Dissertation von Thomas Stuhr stammten (Stuhr, 2004). Damit setzte sich die zu untersuchende Gruppe in Hinblick auf deren ethnische Zugehörigkeit wie folgt zusammen:

6 Asiaten  
3 Aborigenes  
6 Afrikaner  
31 Kaukasier

##### 4.4.1 Theoretische Vorüberlegung

Die flankierende Sequenz, die das HERV-K113-Provirus, wenn es integriert wurde, umrahmt, erscheint erneut stromabwärts in einer duplizierten Version. In dieser Region kommt es allerdings zu keiner Integration von HERV-K113. In den Sequenzanalysen beider Regionen fanden sich zahlreiche Veränderungen in der Basenfolge (Stuhr, 2004). Die flankierende Sequenz, die mit den Primern K113 for und K113 rev erfasst wurde, konnte nur amplifiziert werden, wenn das Provirus nicht integriert war (Abb. 23).

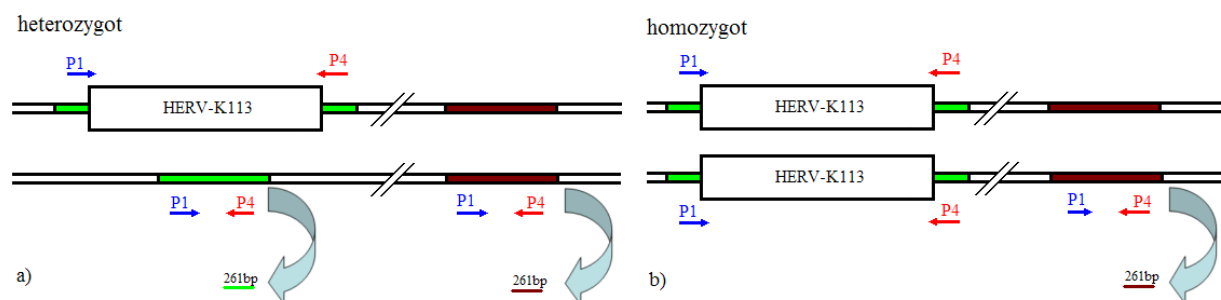


Abb. 23: PCR-Schema zum Nachweis von Homozygotie mit den Primern K113 for (P1) und K113 rev (P4). Bei heterozygot (a) vorliegendem Provirus erhält man zum einen ein Fragment mit 261bp Länge, das die Präintegrationsstelle repräsentiert. Zum anderen entsteht ein Fragment gleicher Länge aus der duplizierten Region. Liegt das Provirus homozygot (b) im Genom der Wirtszelle, amplifizieren die Primer nur die duplizierte Region und es entstehen ebenfalls Fragmente der Länge 261bp

Grund dafür ist, dass bei vorhandenem Provirus der Abstand zwischen den Oligonukleotiden zu groß gewesen wäre, um amplifizieren zu können. Da diese Primer auch in der Lage waren, die duplizierte Region darzustellen, konnte man bei dem PCR-Produkt nicht eindeutig sagen, um welche Region es sich handelte. Vorversuche zeigten, dass die oben genannten Abweichungen nicht genutzt werden konnten, um mit Hilfe der PCR und speziell auf die jeweilige Sequenz zugeschnittenen Primern den Nachweis zu erbringen, ob das Provirus homozygot oder heterozygot vorlag. Somit musste für den Nachweis von Reinerbigkeit eine andere Methode etabliert werden.

#### 4.4.2 Untersuchung der Homo- und Heterozygotie mit Hilfe von Restriktionsendonukleasen

Da es nicht möglich war aufgrund Stringenz gesteuerter PCR-Analysen hetero- von homozygoten Merkmalsträgern zu unterscheiden, wurde versucht, durch selektive enzymatische Restriktion des PCR-Produktes der Primer K113 for und K113 rev nachzuweisen, wie das HERV-K113-Provirus im menschlichen Genom vorlag. Sowohl die Präintegrationsstelle als auch die duplizierte Region wurden in Hinblick auf mögliche Schnittstellen von Endonukleasen analysiert. Anhand der Untersuchung der Sequenzen (Kap.4.2.2) und der Restriktionsmuster verschiedener Enzyme konnte für die flankierende Sequenz die Restriktionsendonuklease *SspI* und für die duplizierte Region das Enzym *AflIII* gefunden werden (Abb.24). Diese Enzyme schnitten ausschließlich und nur einmal in den entsprechenden Sequenzen.

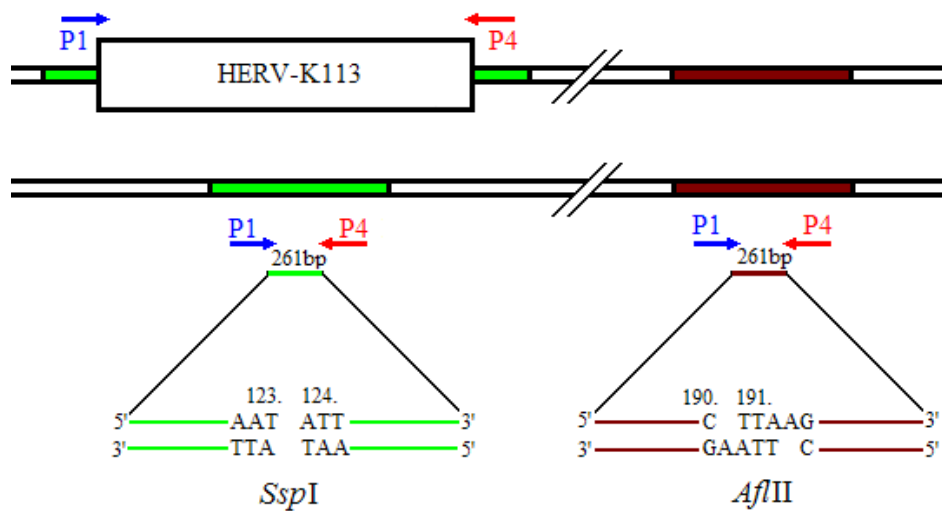


Abb.24: Schematische Darstellung der Restriktionsschnittstellen in den Fragmenten nach PCR mit den Primern K113 for (P1) / K113 rev (P4) für die Restriktionsendonukleasen *Ssp*I in der Präintegrationsstelle und *Afl*III in der duplizierten Region

Das 261bp lange DNA-Fragment wurde in zwei Ansätzen mit jeweils einem der beiden Restriktionsendonukleasen gespalten und nach der Inkubation auf ein 15%iges Polyacrylamidgel aufgetragen. Diese Gele wurden den bisher verwendeten Agarosegelen vorgezogen, da das Auflösungsvermögen im Bereich unterhalb von 100 Basenpaare um ein vielfaches besser und die benötigte DNA-Menge geringer ist. Die Schnittstelle von *Ssp*I lag zwischen der 123. und 124. Base, so dass die Spaltprodukte eine Länge von 123- bzw. 137bp hatten. Da auch das Auflösungsvermögen von Polyacrylamidgelen begrenzt ist, entsteht nach der Spaltung eine breite Doppelbande. *Afl*III schnitt das Fragment an der 190. Base. Damit hatten dessen Fragmente eine Länge von 190 bzw. 70bp. Die Abb.25 zeigt das theoretisch zu erwartende Ergebnis dieses Versuchs. Bei einem homozygoten Merkmalsträger konnte *Ssp*I das Fragment nicht schneiden, da nur die duplizierte Region amplifiziert wird, die keine passende Schnittstelle aufwies. Aufgrund der vorhandenen Schnittstelle für *Afl*III in der duplizierten Region zeigten sich zwei Banden. Lag das Provirus heterozygot vor, bestand das PCR-Produkt erwartungsgemäß aus der duplizierten Region und der flankierenden Sequenz.

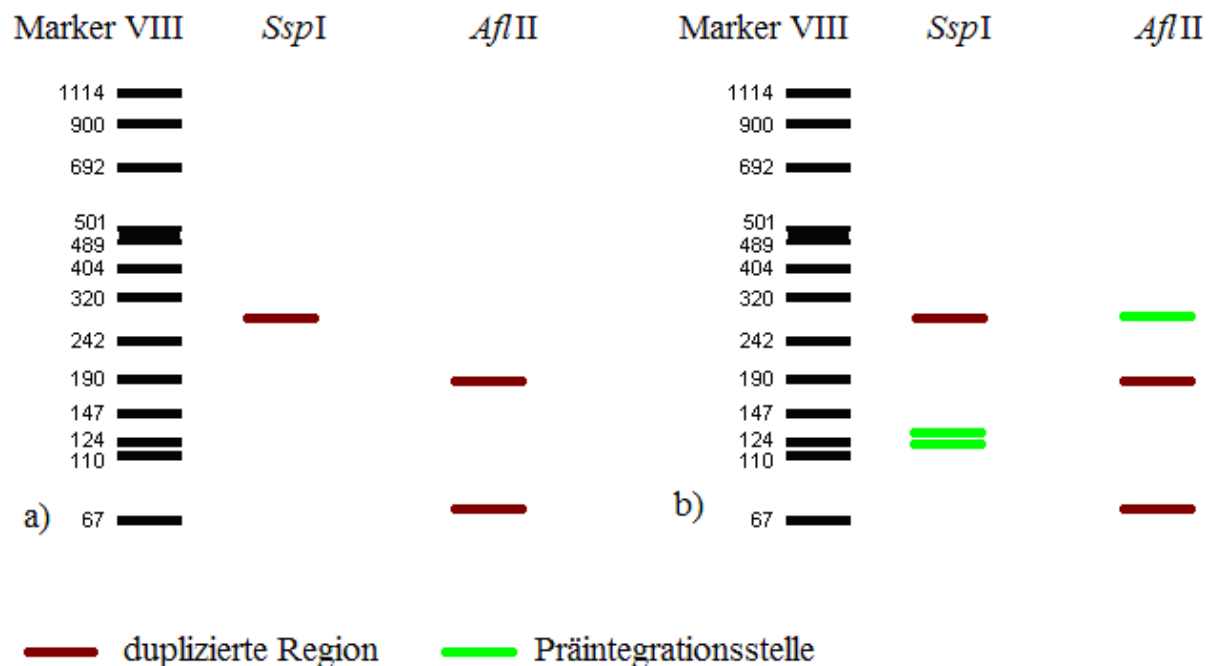


Abb.25: Zu erwartende Banden auf einem 15%Polyacrylamidgel nach Einsatz der Restriktionsendonukleasen *SspI* und *AflII*; a) Ergebnis bei homozygotem Vorliegen des HERV-K113-Provirus; b) Ergebnis bei heterozygotem Vorliegen des HERV-K113-Provirus

Deshalb zeigten sich in diesem Fall zusätzlich zu den bereits erwähnten zwei weitere Banden. Zum einen wurden in dem Ansatz mit *SspI* dessen Spaltprodukte des PCR-Fragments sichtbar. Zum anderen stellte sich ein von *AflII* nicht gespaltenes Fragment dar, bei dem es sich um die flankierende Sequenz handelte. Daraus wird ersichtlich, dass der Nachweis von Homozygotie auch mit einer Restriktionsendonuklease möglich sein sollte. Da es während der Polymerasekettenreaktion jedoch immer zu Punktmutationen innerhalb der amplifizierten Sequenz und damit zu einem Verlust einer der Schnittstellen der Nukleasen kommen konnte, war diese Redundanz aber durchaus sinnvoll.

### 4.4.3 Untersuchung von heterozygoten Merkmalsträgern

Während der Untersuchung stellte sich heraus, dass bei einem heterozygot vorliegenden Provirus die duplizierte Region durch die PCR nicht amplifiziert werden konnte. Das zu erwartende PCR-Produkt hätte sowohl die duplizierte Region als auch die Präintegrationsstelle enthalten müssen. Die Abb.26a zeigt das Ergebnis einer Enzymspaltung mit *Afl*II. Das PCR-Produkt von heterozygoten Merkmalsträgern blieb nach der Inkubation intakt. Die Enzymaktivität von *Afl*II wurde durch Kontrollversuche geprüft (Daten nicht gezeigt). Das erwartete, aus der theoretischen Vorüberlegung folgende Ergebnis ist in Abb.26b dargestellt. Nach dem Verdau sollten Banden mit einer Länge von 191 und 70bp entstehen.

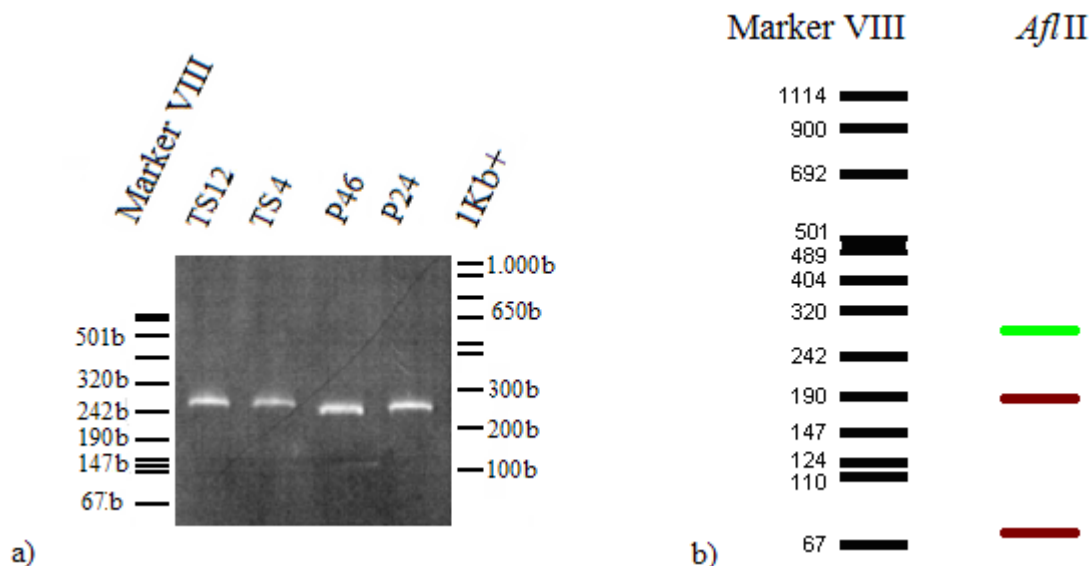


Abb. 26: a) ungeschnittenes PCR-Fragment von vier heterozygoten Merkmalsträgern auf einem 15% Polyacrylamidgel nach Verdau mit *Afl*II; b) zu erwartendes Ergebnis nach Verdau des PCR-Produktes mit *Afl*II, bei dem die duplizierte Region nicht aber die Präintegrationsstelle geschnitten wurde

Dieser Umstand könnte sich aus potentiellen Sequenzunterschieden beider Regionen erklären. Durch einen bisher nicht bekannten Polymorphismus beider Regionen kann nicht ausgeschlossen werden, dass K113 for und K113 rev eine höhere Affinität zu der Präintegrationsstelle als zur duplizierten Region besitzen. Damit wird bei einem heterozygoten Merkmalsträger vorwiegend die flankierende Sequenz amplifiziert.

#### 4.4.4 Untersuchung der Merkmalsträger mit *SspI*

Entsprechend den theoretischen Vorüberlegungen erbrachte der Einsatz der Restriktionsendonuklease *SspI* für homo- und heterozygote Merkmalsträger die Ergebnisse, die zu erwarten waren. Die Abb.27a zeigt diesen Sachverhalt nochmals schematisch, wonach *SspI* aufgrund der vorhandenen Schnittstelle die Präintegrationstelle schneidet, die duplizierte Region jedoch in diesem Ansatz intakt bleibt. Abb.27b gibt analog dazu den in dieser Arbeit erbrachten Nachweis exemplarisch für vier Proben wider. Die eingesetzten PCR-Produkte der Proben TS21 und TS24 mit einer Länge von 261bp blieben dabei unversehrt. Die zu den Proben P46 und P24 gehörenden Fragmente der gleichen Länge wurden durch *SspI* in jeweils zwei Fragmente mit 123 bzw 137bp zerschnitten. Aufgrund des geringen Längenunterschiedes entstand auf dem Polyacrylamidgel nachfolgend eine breite Bande. Daraus kann geschlossen werden, dass das HERV-K113-Provirus in den Proben TS21 und TS24 homozygot (++) und in den Proben P46 und P24 heterozygot (+-) vorlag.

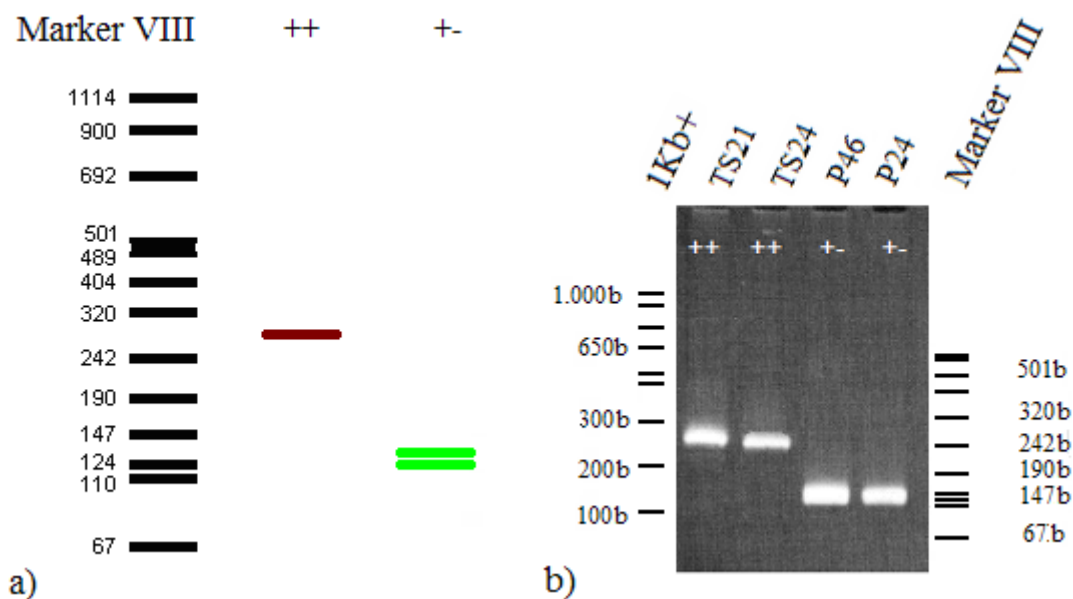


Abb.27: Nachweis von homozygoten (++) und heterozygoten (+-) Merkmalsträgern durch Verdau der mittels PCR gewonnenen Fragmente mit der Restriktionsendonuklease *SspI*; a) schematische Darstellung des erwarteten Ergebnisses mit der intakten **duplizierten Region** und der durch *SspI* geschnittenen **Präintegrationsstelle** b) Ein 15%Polyacrylamidgel zeigt beispielhaft das Ergebnis des Enzymeinsatzes für vier Proben, wobei für TS21 und TS24 Homozygotie und für P46 und P24 Heterozygotie des HERV-K113-Provirus nachgewiesen werden konnte.



#### 4.4.5 Untersuchung der Merkmalsträger mit *Afl*III

Aufgrund der Vorüberlegung und den bisherigen Ergebnissen wurde deutlich, dass der Nachweis für Homozygotie auch mit *Afl*III bestätigt werden kann. In Abb.28b erkennt man das Ergebnis des enzymatischen Verdaus mit *Afl*III für die beiden Proben TS21 und TS24. Es entstanden jeweils Banden mit einer Länge 191 und 70bp. Die Reaktion mit den Proben P46 und P24 konnte in diesem Versuch keine Aktivität von *Afl*III nachweisen. Dementsprechend blieb das eingesetzte Fragment mit einer Länge von 261bp intakt.

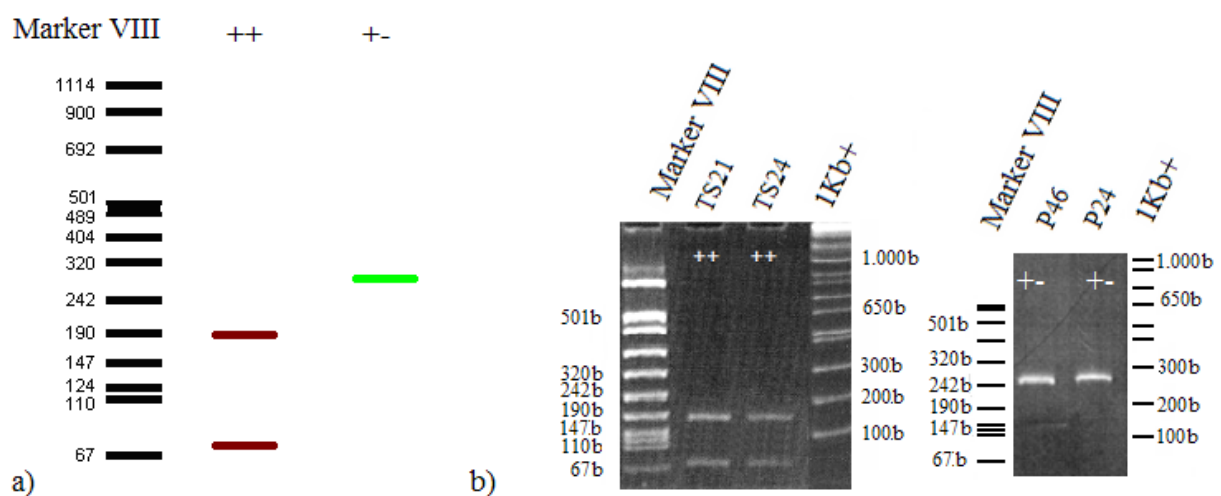


Abb.28: Nachweis von homozygoten (++) und heterozygoten (+-) Merkmalsträgern durch Verdau der mittels PCR gewonnenen Fragmente mit der Restriktionsendonuklease *Afl*III a) schematische Darstellung des erwarteten Ergebnisses mit der durch *Afl*III geschnittenen **duplizierten Region** und der intakten **Präintegrationsstelle** b) Zwei 15%Polyacrylamidgele zeigen beispielhaft das Ergebnis des Enzymsatzes für vier Proben, wobei für TS21 und TS24 Homozygotie und für P46 und P24 Heterozygotie des HERV-K113-Provirus nachgewiesen werden konnte.

Aus den gezeigten Ergebnissen kann geschlossen werden, dass die Proben TS21 und TS24 homzygoten Merkmalsträgern entsprechen, während P46 und P24 heterozygot für das HERV-K113-Provirus sind.

Von den 46 untersuchten DNA-Proben wiesen insgesamt 3 eine homozygote Verteilung des HERV-K113-Provirus auf. Eine stammte aus Afrika, eine aus Australien (TS21) und die dritte hatte eine asiatische Herkunft (TS24).

## 4.5 Sequenzanalyse der duplizierten Region und der Präintegrationsstelle

Die Sequenzen der Präintegrationsstelle und der duplizierten Region sind im Anhang dargestellt (Kap.7.1). Die in dieser Arbeit durchgeführte Analyse ergab innerhalb der Basenfolge beider Regionen 26 Unterschiede.

## 4.6 Sequenzanalyse des *env*- und des *gag*-Gens

Ziel dieser Analyse war es, nach möglichen Sequenzpolymorphismen zwischen den einzelnen Virusgenome zu suchen. Dadurch sollten Rückschlüsse über die Integrität, die potentielle Replikationsfähigkeit und die interindividuelle Varianz der retroviralen Elemente gezogen werden. Mit Hilfe der PCR und Klonierung und Amplifikation in Bakterien wurde das *env*-Gen von sieben und das *gag*-Gen von sechs verschiedenen Merkmalsträgern sequenziert. Als Referenz wurde die jeweilige im Human Genome Project ermittelte Sequenz beider Gene benutzt (Accession No. AY037928). Die vollständigen Sequenzen aller untersuchten Proben sind dem Anhang zu entnehmen.

### 4.6.1 Das *env*-Gen

Die zu untersuchenden Sequenz hat eine Länge von 3026 Nukleotiden. Um diese darzustellen dienten die Oligonukleotide *gesenv* for und K113 rev als Primer für die PCR. Tabelle 12 enthält die beobachteten Sequenzunterschiede. Neun der Nukleotiddifferenzen waren bei allen Proben an derselben Position zu finden. Basierend auf diesen Daten ergibt sich für das *env*-Gen eine Häufigkeitsverteilung für punktuelle Sequenzunterschiede von  $3,63 \times 10^{-3}$ . Das entspricht bei 1000 Nukleotiden einer Änderung der Sequenz von etwa 4 Basen.

Probe	Anzahl der Basenunterschiede	
	Absolut	Relativ
1	12	$4,0 \times 10^{-3}$
2	11	$3,6 \times 10^{-3}$
3	10	$3,3 \times 10^{-3}$
4	12	$4,0 \times 10^{-3}$
5	11	$3,6 \times 10^{-3}$
6	10	$3,3 \times 10^{-3}$
7	11	$3,6 \times 10^{-3}$

Tabelle 12: Ergebnis der Sequenzanalyse des *env*-Leserahmens

#### 4.6.2 Das *gag*-Gen

Aus der PCR mit den Primern K113 for und gesgag rev ging ein Fragment mit einer Länge von 2083 Nukleotiden hervor. Die Anzahl der Abweichungen in der Nukleotidfolge sind Tabelle 13 zu entnehmen, wobei hier ebenfalls erwähnt werden muss, dass fünf Abweichungen bei allen Proben an derselben Stelle zu finden waren. Dem Vorgehen bei *env* entsprechend ergibt sich bei *gag* eine Häufigkeitsverteilung für Unterschiede in der Sequenz von  $3,26 \times 10^{-3}$ . Für dieses Gen entspricht das bei 1000 Nukleotiden einer Abweichung von ca. 3 Basen verglichen mit der Referenzsequenz.

Probe	Anzahl der Basenunterschiede	
	Absolut	Relativ
1	5	$2,4 \times 10^{-3}$
2	9	$4,3 \times 10^{-3}$
3	5	$2,4 \times 10^{-3}$
4	6	$2,9 \times 10^{-3}$
5	9	$4,3 \times 10^{-3}$

Tabelle13: Ergebnis der Sequenzanalyse des *gag*-Leserahmens

## 5. Diskussion

Mit den Arbeiten von Barbulescu et al. (1999) und Turner et al. (2001) wurden erstmals Daten veröffentlicht, die sich mit den retroviralen Elementen HERV-K113 und HERV-K115 eingehend beschäftigten und diese in den bis dato bestehenden wissenschaftlichen Kontext einordneten. Zusammen stellen sie den Ausgangspunkt der vorliegenden Arbeit dar. Die sich daraus ergebenden Fragestellungen werden nochmals kurz umrissen und die gewonnenen Daten unter Berücksichtigung der aktuellen Literatur diskutiert.

### 5.1 Fragestellungen der Arbeit

Die Frage, wie häufig beide endogene Retroviren in der menschlichen Population vorkommen, bildet den Kern dieser Arbeit. Da das von Turner et al. 2001 untersuchte Kollektiv zum einen ethnisch sehr heterogen war und zum anderen nur aus 31 Probanden bestand, war es das Ziel, eine größere, ethnisch weitestgehend homogene Gruppe auf das Vorhandensein beider Proviren zu untersuchen. Einige Mitglieder der HERV-K-Familie werden mit der Entstehung von keimbahnassoziierten Tumoren in Verbindung gebracht. Um die Frage zu klären, in wie weit dies auch für die hier analysierten Retroviren gilt, bestand ein Teil des Kollektivs aus Patienten mit eben solchen Keimzelltumoren.

Eine andere Ebene dieser Arbeit ergibt sich ebenfalls aus den genannten Veröffentlichungen. Turner und Mitarbeiter entwickelten ein PCR-Schema um nachzuweisen, ob die Proviren in homo- oder heterozygoter Form im Genom der Merkmalsträger vorlagen. Mit Hilfe dieser Untersuchung lassen sich Rückschlüsse auf den Zeitpunkt und den Ort der Erstintegration der untersuchten Retroviren ziehen. Die vorliegende Arbeit beschränkt sich mit der Analyse auf das humane endogene Retrovirus K113.

Da es sich bei diesem im Gegensatz zu HERV-K115 um ein Volllängenprovirus mit offenen Leserahmen für alle retroviralen Proteine handelt, ist die Sequenzanalyse des *env*- und *gag*-Gens wiederum nur bei HERV-K113 von Interesse. Der Vergleich sequenzierter Gene mehrerer Individuen mit einer Referenzsequenz sollte Aufschluss über die Integrität des proviralen Genoms bringen und etwaige Unterschiede in der Mutationshäufigkeit aufzeigen.

## 5.2 Genfrequenz und Bedeutung von HERV-K113 und HERV-K115 für den Menschen

Die in dieser Arbeit gewonnenen Zahlen zeigen deutlich, dass die von Turner et al. (2001) publizierten Daten bezüglich der Häufigkeit von HERV-K113 und HERV-K115 für eine europäisch/kaukasische Population nicht einfach übernommen werden können. Bei genauerer Betrachtung der erwähnten Arbeit stellt man fest, dass das untersuchte Kollektiv mit Individuen aus Afrika, Amerika, Europa, Ozeanien und Ostasien einerseits sehr heterogen zusammengesetzt und die Fallzahl andererseits mit 31 Proben gering war. Die Insertionspolymorphismen beider endogener Retroviren mit Werten von 29% für HERV-K113 und 16% für HERV-K115 sind demzufolge eher als globaler Trend zu werten, der nicht beliebig auf verschiedene Regionen der Welt übertragen werden kann. Mit einer Anzahl von 207 Proben aus dem europäisch/kaukasischen Raum und einer Merkmalsverteilung von 13,6% für HERV-K113 und 4,4% für HERV-K115 zeigen die Daten der vorliegenden Arbeit, mit welcher Frequenz beide Proviren in der untersuchten Region vorkommen. Allerdings ist aufgrund der immer noch geringen Fallzahl Vorsicht bei der Interpretation der Ergebnisse geboten. Moyes et al. veröffentlichten 2005 eine Arbeit, die sich zum Teil ebenfalls mit diesem Thema beschäftigte. Gegenstand der Untersuchung war hier die Suche nach HERV-K113 und HERV-K115 mit Hilfe eines PCR-Schemas in Kollektiven aus Afrika, Großbritannien, dem Jemen und Papua Neu Guinea. Es zeigte sich für die afrikanische Population eine Häufigkeit von 21,8 bzw. 34,1%. Die Prävalenz in Großbritannien lag lediglich bei 4,2 bzw. 1%, die im Jemen bei 8 und 7,1%. In den 54 Proben aus Papua Neu Guinea waren beide Proviren nicht nachzuweisen. Damit ist klar, dass das Auftreten beider retroviralen Elemente direkt mit der ethnischen Herkunft des untersuchten Individuums zusammenhängt. Eine Arbeit von Herrera et al. untermauert diese These nochmals (Herrera et al., 2006). Diese Arbeitsgruppe untersuchte DNA-Proben von Menschen aus Afrika, dem mittleren Osten, Südwesteuropa und Ostasien. Hervorzuheben ist hier, dass, wie auch bei Moyes et al., innerhalb der einzelnen Regionen die Populationen nach Ländern geordnet waren und die afrikanischen Kollektive aus Ländern südlich der Sahara stammten. Für HERV-K113 zeigte sich die höchste Prävalenz mit 35,9% in Kenia und die niedrigste mit 2% in Nordspanien. Die Gesamtprävalenz lag bei 12,2%. Für HERV-K115 lag diese bei 12,7%, wobei der tiefste Wert mit 4% in China und der höchste mit 22,4% in Ruanda zu finden war. Bemerkenswert an diesen Zahlen ist aber auch, dass sich das Verhältnis der

Proviren zueinander ebenfalls regional zu verändern scheint. Während in Kenia und Ostasien die Häufigkeit von HERV-K113 über der von HERV-K115 liegt, kehrt sich dieses Verhältnis im mittleren Osten, in Kamerun und auch in Südwesteuropa um. Allerdings sind auch hier die Zahlen zu klein, um eindeutige Aussagen machen zu können. In Kooperation mit der Arbeitsgruppe von Prof. Dr. Müller-Lantzsch untersuchten Moussa et al. 129 Individuen pygmäischer Herkunft in Zentralafrika auf das Vorhandensein beider Proviren (unveröffentlichte Daten). Dabei fanden sie mit 75% eine sehr hohe Prävalenz für HERV-K113, wohingegen die Prävalenz von HERV-K115 bei 10% lag. Der hohe Anteil an Merkmalsträgern lässt sich wahrscheinlich durch zwei Phänomene erklären. Zum einen wurden bei der erwähnten Untersuchung ganze Familien analysiert womit es zu einer Verzerrung der Daten kommt. Zum anderen handelt es sich bei den Pygmäen um eine kleine, in sich geschlossene Population, in der Genaustausch mit anderen Volksgruppen weniger häufig stattfindet, als in anderen Kulturkreisen. In der Zusammenschau all der bisher erhobenen Daten kann man allerdings erkennen, dass die Prävalenz beider Proviren in afrikanischen Ländern südlich der Sahara mit ungefähr 30% für HERV-K113 und 20% für HERV-K115 weltweit am höchsten ist. Je weiter man sich von dieser Region entfernt, desto niedriger scheint die Prävalenz zu werden. Diese Beobachtung stützt die Hypothese, der Mensch hätte sich in Afrika entwickelt und sich dann vor ca. 150 -200000 Jahren nord- und ostwärts ausgebreitet (Jin et al., 2000). Dieser Zeitraum korreliert auch mit dem Zeitpunkt, den Turner et al. (2001) für die Erstintegration von HERV-K113 in die menschlichen Keimbahnzellen festgelegt haben. Eine Infektion während oder nach der Migration würde erklären, warum die Prävalenz in Afrika, verglichen mit dem Rest der Welt, erhöht ist.

Der bloße Vergleich der Prävalenzen innerhalb verschiedener ethnischer Gruppen und die daraus resultierenden Hinweisen auf den Ursprung des *Homo sapiens*, ist ein Aspekt. Die Suche nach einem möglichen Zusammenhang zwischen den analysierten Proviren und Erkrankungen ein anderer. Vor allem die Krankheiten, die bereits mit anderen Mitgliedern der HERV-K-Familie in Verbindung gebracht wurden, stehen im Fokus. Neben dem Seminom, das Gegenstand der vorliegenden Arbeit ist, haben sich andere Arbeitsgruppe mit der Beziehung von HERV-K113 und HERV-K115 zu Autoimmunerkrankungen (Moyes et al., 2005), zu Mammakarzinomen (Burmeister et al., 2004) und zu psychiatrischen Erkrankungen (Otowa et al., 2006) auseinandergesetzt.

Dass es einen Zusammenhang zwischen der Tumorigenese des Seminoms und retroviralen Elementen der HERV-K-Familie gibt, kann mittlerweile als gesichert angesehen werden. Die konkrete Frage, ob diese Beziehung auch für die untersuchten Elemente gilt, beantwortet die

vorliegende Arbeit nur teilweise. Für HERV-K115 ist eine direkte Korrelation von Tumor und Provirus eher unwahrscheinlich, da sich die Prävalenzen in Kontroll- und Seminomgruppe nicht wesentlich unterscheiden. Ein Gesichtspunkt, den diese Arbeit nicht berücksichtigt, ist der Malignitätsgrad bzw. das Staging der Hodentumore in der untersuchten Population. Somit kann nicht gesagt werden, ob ein Zusammenhang zwischen dem Vorhandensein des Provirus und der Aggressivität des Seminoms besteht. Dies sollte Gegenstand weiterer Untersuchungen sein. Die Unterschiede der Häufigkeitsverteilung in beiden Gruppen bezogen auf HERV-K113 lassen mehr Raum für Spekulationen. Die Daten bezüglich der Genfrequenz bzw. der Prävalenzen - 0,09 bzw. 18,8% in der Kontrollgruppe und 0,06 bzw. 11,1% in der Seminomgruppe – geben der Spekulation Raum, das Provirus habe bezüglich der Entstehung von Seminomen protektive Eigenschaften für den Merkmalsträger. Dass Endogene Retroviren ihrem Wirt nicht nur schaden, haben verschiedene Arbeiten bereits gezeigt. So könnte die Integration des Provirus auf Chromosom 19 in der Nähe eines Tumorsuppressorgens stattgefunden haben, das dadurch in seiner Aktivität gefördert wird. Ein anderes Denkmodell beruht auf dem Prinzip des molekularen Mimikry. HERV-K113 mit seinen intakten offenen Leserahmen für alle retroviralen Proteine scheint ein replikationsfähiges Virus zu sein. Wenn dem so ist, reagiert das Immunsystem des Wirtsorganismus mit der Bildung von Antikörpern gegen diese Proteine. Theoretisch wäre es möglich, dass die Oberflächen der entarteten Seminomzellen denen der viralen Proteine gleichen und so einen Angriffspunkt für die humorale Immunantwort bilden. Eine Arbeit, die Herve et al. 2002 veröffentlichten, stützt diese These. Darin zeigten die Autoren, dass das Immunsystem von Menschen, in deren Genom sich Proviren der HERV-K-Familie integriert hatten, mit der Bildung von spezifischen Immunglobulinen reagierte. Gegenstand der Forschung waren hier allerdings nicht Keimzelltumoren, sondern Autoimmunerkrankungen aus dem rheumatischen Formenkreis. Die bereits erwähnte Studie von Moyes et al. (2005) beschäftigt sich eingehend mit dem Zusammenhang solcher Krankheiten und dem Auftreten von HERV-K113 und HERV-K115. Im Fokus standen hier die Multiple Sklerose (MS), das Sjögren Syndrom (SS) und die rheumatoide Arthritis (RA). Einer Kontrollgruppe von jeweils 96 Personen aus Großbritannien wurden im Fall von HERV-K113 109 Patienten mit MS, 96 mit SS und 96 mit RA gegenübergestellt. Bei HERV-K115 waren es 92 Patienten mit MS und jeweils 96 mit SS und RA. Während für letzteres Provirus keine signifikante Steigerung der Prävalenz in einer der Patientenkollektive zu verzeichnen war, zeigte sich eine statistisch signifikante Erhöhung für HERV-K113 in der MS- ( $p < 0,05$ ) und der SS-Gruppe ( $p < 0,01$ ). Mit 11,9 bzw. 15,9% lag die Verteilung deutlich über den 4,2% in der Kontrollgruppe. Damit

schließt sich dieses Provirus einer wachsenden Gruppe von endogenen Retroviren an, die mit Multipler Sklerose in Verbindung gebracht werden (Rasmussen et al., 1995; Clerici et al., 1999). Die Autoren weisen allerdings darauf hin, dass zum einen die Fallzahl insgesamt gering war. Zum anderen geben sie zu bedenken, dass die Zusammensetzung des SS-Kollektivs ethnisch sehr heterogen war und dadurch eine Verzerrung der Daten möglich sei. Da sich das Verhältnis von HERV-K113 zu HERV-K115, das in der Kontrollgruppe nahezu gleich war, in der Gruppe mit Sjögren Syndrom stark zu Gunsten von HERV-K113 verschoben hatte, gehen Moyes et al. aber von validen Daten aus. Neben den exogenen Retroviren HTLV-1 und HIV-1, die mit der Entstehung des Sjögren Syndrom in Verbindung gebracht werden (Venables, 2004), scheint ein Zusammenhang von HERV-K113 mit diesem Symptomkomplex möglich.

Ausgehend von den Arbeiten von Wang-Johanning et al., die sich mit der Beziehung von HERV-K und der Entstehung von Mammakarzinomen beschäftigen, geht eine Studie von Burmeister et al. näher auf HERV-K113 und HERV-K115 ein (Burmeister et al., 2004). Darin vergleicht die Arbeitsgruppe die Prävalenz beider Proviren bei 102 Patientinnen, die an Brustkrebs erkrankt waren, mit einer gleich großen Kontrollgruppe. Für beide Retroviren konnte keine signifikante Steigerung oder Verminderung der Häufigkeitsverteilung festgestellt werden. In der Kontrollgruppe fanden sich 13 (=12,7%), in der Karzinomgruppe 15 (=15,7%) Personen, die das HERV-K113-Provirus in sich trugen. Für HERV-K115 lag das Verhältnis bei 5 (=4,9%) positiven Merkmalsträgern in der Kontroll- und bei 10 (=9,8%) in der Karzinomgruppe. Auch unter Berücksichtigung des Malignitätsgrades und des Tumorstaging stellten die Autoren keine Änderung der erhobenen Daten fest, so dass ein Zusammenhang zwischen der Entstehung des Mammakarzinoms und den endogenen Retroviren HERV-K113 und HERV-K115 als wenig wahrscheinlich gilt.

Vermehrte Hinweise, dass sowohl exogene (Lewis, 2001) als auch endogene Retroviren (Yolken et al., 2000; Frank et al., 2005) an der Entstehung von Schizophrenie und schizoaffektiven Störungen beteiligt sein könnten, bilden die Grundlage für die Arbeit von Otowa et al. In dieser Studie untersuchten sie den Zusammenhang zwischen HERV-K115 und dem Auftreten der genannten psychiatrischen Erkrankungen. Dabei fanden sie in einer japanischen Population keine signifikante Korrelation zwischen dem Provirus und der Krankheitsprävalenz. Die Kontrollgruppe lag mit 9,4% in etwa auf gleichem Niveau wie die untersuchte Population (8,4%). Allerdings konnten die Autoren einen schwach signifikanten Zusammenhang zwischen dem Vorhandensein des Provirus und dem Verlauf der Krankheit herstellen. In der Gruppe der early onset schizophrenia, also dem frühzeitig beginnenden



schweren Verlauf der Krankheit, kam das retrovirale Element signifikant häufiger vor ( $p=0,057$ ). Somit hat das Retrovirus wahrscheinlich keinen ätiologischen Effekt, könnte aber durchaus den Verlauf der Erkrankung beeinflussen.

Zusammen mit der existierenden Literatur können die in der vorliegenden Arbeit erhobenen Daten zwar keine Beweise dafür liefern, ob eines der untersuchten retroviralen Elemente in einem konsekutiven Zusammenhang zu der Entstehung einer Erkrankung, sei sie autoimmuner oder maligner Genese, steht. Sie können aber Hinweise dafür geben, ob eine Beteiligung der HERV als wahrscheinlich gilt und damit eine fokussierte Forschung ermöglichen.

### 5.3 Homozygotie des humanen endogenen Retrovirus K113

In den bereits besprochenen Veröffentlichungen benutzen die jeweiligen Arbeitsgruppen ein PCR-Schema um nachzuweisen, ob das Provirus homozygot vorliegt. Ausgehend von dieser Methode beschrieben Turner et al. in ihrer Arbeit von 2001 zwei Merkmalsträger mit homozygoter Verteilung für HERV-K113. Beide stammen aus afrikanischen Regionen südlich der Sahara. Moyes et al. (2005) fanden unter den 374 von ihnen untersuchten Menschen vier für HERV-K113 homozygote Individuen. Auch diese stammen aus afrikanischen Ländern südlich der Sahara. Burmeister et al. (2004) wiesen für keine der von ihnen untersuchten Patientinnen ein reinerbig vorliegendes Provirus nach. In dieser Arbeit wurde eine Methode, beruhend auf Restriktionsendonukleasen, etabliert, die eine sichere Aussage über Rein- oder Mischerbigkeit der untersuchten Merkmalsträger erlaubt. Unabhängig von möglichen Fehlhybridisierungen während der Polymerasekettenreaktion, ermöglicht der Einsatz von Restriktionsendonukleasen einen reproduzierbaren Nachweis von Homozygotie des HERV-K113-Provirus. Unter den 46 positiven DNA-Proben, bei denen das HERV-K113-Provirus nachgewiesen wurde, fanden sich drei homozygote Merkmalsträger. Einer von ihnen stammt aus Afrika, der andere aus der australischen Urbevölkerung. Die dritte reinerbige Merkmalsträger war asiatischer Herkunft. Damit bestätigen diese Daten die bisher veröffentlichten nur teilweise. Sowohl Moyes et al. (2005) als auch Turner et al. (2001) konnten Reinerbigkeit jeweils nur bei aus Afrika stammenden Personen nachweisen. Auch Burmeister et al. (2004) beschreiben lediglich heterozygote Merkmalsträger in der untersuchten hellhäutigen Population. Dass eine der homozygoten Personen dieser Arbeit aus Afrika stammt, untermauert die These, die Erstinfektion hätte vor etwa 150 - 200000 Jahren

in einer südlich der Sahara liegenden Region Afrikas stattgefunden. Da Moyes und Mitarbeiter in ihrer Population aus Ozeanien nicht ein Provirus nachweisen konnten, stellten sie die Hypothese auf, die Integration hätte möglicherweise erst vor 50 - 60000 Jahren stattgefunden. Dass in der vorliegenden Arbeit eine der reinerbigen Personen aus der australischen Urbevölkerung stammt und eine andere aus dem asiatischen Raum, widerlegt diese Annahme zwar nicht, schwächt sie aber. Da aber sowohl in der vorliegenden Arbeit als auch in der von Moyes et al. (2005) die Fallzahl sehr gering war, steht eine abschließende Bewertung noch aus.

Dass eine Häufung der Reinerbigkeit und der Prävalenz des Provirus in Afrika zu beobachten ist, stützt die Hypothese, der Ursprung des heutigen Menschen läge auf diesem Kontinent. Eine andere mögliche Erklärung basiert auf Selektionsdruck. Als Beispiel für einen solchen Effekt sei die Sichelzellanämie genannt. Die Prävalenz dieser Erkrankung ist in Malariaendemiegebieten Afrikas signifikant erhöht, da heterozygote Merkmalsträger durch einen leichteren Verlauf der Malaria tropica eine relative Resistenz entwickeln. Außerhalb Afrikas übt diese Erkrankung einen negativen Selektionsdruck auf dortige Populationen aus und ist dementsprechend selten anzutreffen. Danach könnte die Integration von HERV-K113 die Menschen in Afrika vor der Infektion durch exogene Retroviren geschützt haben, in Eurasien aber eventuell zu Autoimmun- oder Krebserkrankungen führen. Betrachtet man die Inzidenz für Keimzelltumoren in den verschiedenen Entitäten, kann man ein deutliches Nord-Süd-Gefälle erkennen (Schmelz und Albers, 2007). Die geographische Verteilung zeigt die höchste Inzidenz in Skandinavien, Deutschland und Neuseeland, ein mittleres Vorkommen in den USA. Unter Schwarzafrikanern und Asiaten sind Keimzelltumoren sehr selten. Diese Verteilung korreliert deutlich mit den in den diskutierten Untersuchungen und der vorliegenden Arbeit erhobenen Daten.

## 5.4           HERV-K 113 und seine provirale Integrität

Das *env*-Gen nimmt innerhalb des proviralen Genoms der humanen endogenen Retroviren eine Sonderstellung ein. Die mit diesem Leserahmen assoziierten Genprodukte bilden den Teil eines Retrovirus, auf den das spezifische Immunsystem des Wirtsorganismus durch humorale und zelluläre Antwort schnell und effektiv reagieren kann. Damit ist der Nachweis von Antikörpern gegen Env-Proteine zum einen ein indirekter Hinweis für das Vorhandensein eines retroviralen Elements. Zum anderen bildet ein intakter Leserahmen die Grundlage für

eine funktionsfähige Transkription und Proteinbiosynthese. Mehrere Arbeitsgruppen haben zeigen können, dass die Integration verschiedener Retroviren der HERV-K-Familie mit einem hohen Titer an Antikörpern assoziiert ist, die gegen Env-Proteine gerichtet sind (Sauter et al., 1996; Herve et al., 2002). Als direkte Bestätigung für eine vollständige Genexpression kann man entweder einzelne Proteine oder das daraus resultierende Produkt nachweisen. Im Fall von *env* gelang beides. Zum einen konnte gezeigt werden, dass es sich bei dem humanen, mit Teratokarzinomen assoziierten Virus (HTDV), das virus-ähnliche Partikel produziert, um HERV-K handelt. Zum anderen ist die Genexpression im *env*-Leserahmen von HERV-K durch In-situ-Hybridisierung belegt. Neben dem für die Virushülle verantwortlichen Teil des retroviralen Genoms, spielt das *gag*-Gen für die Bildung eines Viruspartikels die entscheidende Rolle. Für HERV-K10 ist belegt, dass dieses Gen intakt und zur Bildung eines funktionsfähigen Proteins fähig ist (Müller-Lantzsch et al., 1993). Diese Arbeitsgruppe sequenzierte das Gen und stellte basierend auf der Nukleotidfolge das entsprechende Protein her. Den bisher bei anderen Mitglieder der HERV-K-Familie durchgeführten Analysen entsprechend, sollte die Sequenzierung des *env*- und des *gag*-Gens in der vorliegenden Arbeit Aufschluss über die provirale Integrität von HERV-K113 geben. Die Daten zeigen, dass innerhalb der Leserahmen beider Gene keine Deletionen oder Rasterverschiebungen vorliegen. Die nachgewiesenen Unterschiede in der Basenfolge liegen mit 3-4 pro 1000 Nukleotiden deutlich unterhalb dessen, was als signifikant zu bezeichnen wäre. Man kann also von einer Sequenzhomologie im Bereich des *env*- und des *gag*-Leserahmens innerhalb der untersuchten Population mit einem unauffälligen Polymorphismus ausgehen. Damit lassen diese Ergebnisse den Schluss zu, dass das humane endogene Retrovirus K113 auf proviraler Ebene die Möglichkeit besitzt, Viruspartikel zu synthetisieren, die, mit einem kompletten retroviralen Genom versehen, zur Reinfektion befähigt sind. Dass die Mitglieder der HERV-K-Familie theoretisch dazu in der Lage sind, konnten Dewannieux et al. 2006 zeigen. Aus neun für Menschen spezifischen Mitgliedern von HERV-K(HML-2) generierten sie mit HERV-K108 und HERV-K109 als Rückgrat ein neues Humanes Endogenes Retrovirus, *Phoenix*. Es unterscheidet sich in seiner Sequenz in mindestens 20 Aminosäuren von den Referenzelementen und weist in seinem 5'-LTR in der U3-Region einen CMV-Promotor auf. Diese Variation sollte eine sichere Transkription gewährleisten, da diese Region nicht in allen Zelllinien funktionsfähig ist. Dieses so kreierte retrovirale Element wurde im Folgenden als Modell benutzt, um die Amplifikations-, Infektions- und Rekombinationsvorgänge von HERV-K(HML-2) nachvollziehen zu können. Tatsächlich konnte die Arbeitsgruppe mit Hilfe von Elektronenmikroskopie die Bildung, Knospung und extrazelluläre Reifung von

Viruspartikeln zeigen. Die Proteinbiosynthese von Env und Gag, die diesem Schritt vorausging, konnte durch Immunoblot-Analyse gezeigt werden. Auch die Reinfektion von menschlichen Zellen gelang *in vitro*. Die Integration des Provirus lief dabei nach dem immer gleichen Muster ab, wobei keine eindeutige, auf den Ort bezogene Präferenz nachweisbar war. Darüber hinaus stellten sie sich die Frage, wie die intraindividuelle Vermehrung von *Phoenix* von statten ging. Die diesbezüglich durchgeführten Versuche und daraus hervorgehenden Daten untermauerten die Hypothese, die Amplifikation von HERV-K(HML-2) finde via Reinfektion und nicht durch intrazelluläre Retrotransposition statt. In wie weit der Prozess der Rekombination Einfluss auf die Aktivität von humanspezifischen Endogenen Retroviren nehmen kann, sollte ein anderer Ansatz der Arbeit klären. Die Arbeitsgruppe um Dewannieux generierten Chimeren aus HERV-K108, -K109, -K113 und -K115 und stellte damit mögliche chromosomale Rekombinationsvorgänge *in vitro* nach. Und tatsächlich konnten sie für die Chimere HERV-K109/-K115 Transkriptionsaktivität nachweisen. Da HERV-K115 ein defektes env-Gen aufweist, erstellten sie desweiteren eine Dreier-Chimere aus HERV-K109, -K115 und -K108 und wiesen für dieses rekombinierte endogene Retrovirus mit Hilfe der quantitativen PCR ein eindeutiges Infektionspotential nach. Dass Rekombination in diesem Ausmass innerhalb Endogener Retroviren stattfindet, konnte bereits gezeigt werden (Hughes und Coffin, 2005). Es liegt also durchaus im Bereich des möglichen, dass Mitglieder der HERV-K-Familie zusammen mit funktionsfähigen Elementen des Wirtsgenoms, eigene funktionsunfähige Regionen reaktivieren und dadurch ihren Lebenszyklus fortsetzen können.

## 6. Literaturverzeichnis

Andersson ML, Lindeskog M, Medstrand P, Westley B, May F, Blomberg J (1999) Diversity of human endogenous retrovirus class II-like sequences. *J. Gen. Virol.*; 80:255-260

Armbruester V, Sauter M, Krautkraemer E, Meese E, Kleiman A, Best B, Roemer K, Mueller-Lantzsch N (2002) A novel gene from the human endogenous retrovirus K expressed in transformed cells. *Clin Cancer Res.*; 8:1800-1807

Armbruester V, Sauter M, Roemer K, Best B, Hahn S, Nty A, Schmid A, Philipp S, Mueller A, Mueller-Lantzsch N (2004) Np9 protein of human endogenous retrovirus K interacts with ligand of numb protein X. *J Virol.*; 78:10310-10319

Baltimore D (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature*; 226:1209-11

Barbulescu M, Turner G, Seaman MI, Deinard AS, Kidd KK, Lenz J (1999) Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans. *Curr. Biol.*; 9:861-868

Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dautet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, Rozenbaum W, Montagnier L. (2004) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). 1983. *Rev Invest Clin.* 56:126-9.

Belshaw R, Dawson AL, Woolven-Allen J, Redding J, Burt A, Tristem M (2005) Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): implications for present-day activity. *J Virol.*; 79:12507-12514

Bishop JM (1991) Molecular themes in oncogenesis. *Cell*; 64:235-248

Bittner JJ (1936) Some possible effects of nursing on the mammary gland tumour incidence in mice. *Science*; 84:162-162

Blond JL, Beseme F, Duret L, Bouton O, Bedin F, Perron H, Mandrand B, Mallet F (1999) Molecular characterization and placental expression of HERV-W, a new human endogenous retrovirus family. *J Virol.*; 73:1175-1185

Boese A, Sauter M, Galli U, Best B, Herbst H, Mayer J, Kremmer E, Römer K, Müller-Lantzsch N (2000) Human endogenous retrovirus protein cORF supports cell transformation and associates with the promyelocytic leukemia zinc finger protein. *Oncogene*; 19:4328-4336

Boller K, König H, Sauter M, Müller-Lantzsch N, Löwer R, Löwer J, Kurth R (1993) Evidence that HERV-K is the endogenous retrovirus sequence that codes for the human teratocarcinoma-derived retrovirus HTDV. *Virology*; 196:349-353

Burmeister T, Ebert AD, Pritze W, Loddenkemper C, Schwartz S, Thiel E (2004) Insertional polymorphisms of endogenous HERV-K113 and HERV-K115 retroviruses in breast cancer patients and age-matched controls. *AIDS Res Hum Retroviruses*; 20:1223-1299

Cohen SN, Chang AC (1974) A method for selective cloning of eukaryotic DNA fragments in *Escherichia coli* by repeated transformation. *Mol Gen Genet.*; 134:133-141

Clerici M, Fusi ML, Caputo D, Guerini FR, Trabattoni D, Salvaggio A, Cazzullo CL, Arienti D, Villa ML, Urnovitz HB, Ferrante P (1999) Immune responses to antigens of human endogenous retroviruses in patients with acute or stable multiple sclerosis. *J Neuroimmunol.* ; 99:173-182

Coffin JM (1992) Structure and classification of retroviruses. In: Levy JA (eds) *The retroviridae*, Plenum Press, New York, pp. 19 –49

Coffin, JM; Hughes SH; Varmus HE (eds) (1997) *Retroviruses*. Cold Spring Harbor Laboratory Press, Plainview New York

Dalglish AG, Beverley PC, Clapham PR, Crawford DH, Greaves MF, Weiss RA (1994) The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature*; 312:763-767

Dangel AW, Baker BJ, Mendoza AR, Yu CY (1995) Complement component C4 gene intron 9 as a phylogenetic marker for primates: long terminal repeats of the endogenous retrovirus ERV-K(C4) are a molecular clock of evolution. *Immunogenetics*.; 42:41-52

Dewannieux M, Harper F, Richaud A, Letzelter C, Ribet D, Pierron G, Heidmann T (2006) Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. *Genome Res.*; 16:1548-1556

Ellermann V, Bang O (1908) Experimentelle Leukämie bei Hühnern. *Zentralbl. Bakteriол. Parasitenkd. Infektionskr. Hyg. Abt. Orig.*; 46:595-609

Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA (eds.) (2005) *Eighth Report of the International Committee on Taxonomy of Viruses*, Academic Press London, San Diego

Frank O, Giehl M, Zheng C, Hehlmann R, Leib-Mösch C, Seifarth W (2005) Human endogenous retrovirus expression profiles in samples from brains of patients with schizophrenia and bipolar disorders. *J Virol.*; 79:10890-10901

Furuichi Y, Shatkin AJ, Stavnezer E, Bishop JM (1975) Blocked, methylated 5' terminal sequence in avian sarcoma virus RNA. *Nature*; 257:618-620

Galli UM, Sauter M, Lecher B, Maurer S, Herbst H, Römer K, Müller-Lantzsch N (2005) Human endogenous retrovirus rec interferes with germ cell development in mice and may cause carcinoma in situ, the predecessor lesion of germ cell tumors. *Oncogene*; 24:3223-3228

Gardner MB, Henderson BE, Officer JE, Rongey RW, Parker JC, Oliver C, Estes JD, Huebner RJ (1973) A spontaneous low motor neuron disease apparently caused by indigenous type-C RNA virus in wild mice. *J. Natl. Cancer Inst.*; 51: 1243-1254

Gillespie D, Marshall S, Gallo RC (1972) RNA of RNA tumour viruses contains poly A. *Nat. New Biol.*; 236: 227-231

Herrera RJ, Lowery RK, Alfonso A, McDonald JF, Luis JR (2006) Ancient retroviral insertions among human populations. *J Hum Genet.*; 51:353-62

Herve CA, Lugli EB, Brand A, Griffiths DJ, Venables PJ (2002) Autoantibodies to human endogenous retrovirus-K are frequently detected in health and disease and react with multiple epitopes. *Clin Exp Immunol.*; 128:75-82

Holmes DS, Quigley M, (1981) A rapid boiling method for the preparation of bacterial plasmids. *Anal Biochem.*; 114:193-197

Hughes JF, Coffin JM (2001) Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. *Nat Genet.*; 29:487-489

Hughes JF, Coffin JM (2004) Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: implications for human and viral evolution. *Proc Natl Acad Sci USA*; 101:1668-1672

Hughes JF, Coffin JM (2005) Human endogenous retroviral elements as indicators of ectopic recombination events in the primate genome. *Genetics*; 12:12

Innis MA, Myambo KB, Gelfand DH, Brow MA, (1988) DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc Natl Acad Sci U S A.*; 85:9436-9440

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature*; 409:860-921

Jacobsen S (2002) Immunopathogenesis of human T cell lymphotropic virus type-1-associated neurological disease. *J. Infect. Dis.*; 186:187

Jin L, Su B (2000) Natives or immigrants: modern human origin in east Asia. *Nat Rev Genet.*; 1:126-33

Johnson WE, Coffin JM (1999) Constructing primate phylogenies from ancient retrovirus sequences. *Proc. Natl. Acad. Sci*; 96:10254-10260



Kiermer V, Van Lint C, Briclet D, Vanhulle C, Kettmann R, Verdin E, Burny A, Droogmans L (1998) An interferon regulatory factor binding site in the U5 region of the bovine leukemia virus Long Terminal Repeat Stimulates Tax-Independent Gene Expression *J Virol.*; 72: 5526–5534

Le Pecq JB, Paoletti C (1965) Interaction of ethidium hydrobromate (EH) with polyribonucleotides. Applications to the study of hybridization reactions. *C R Acad Sci Hebd Seances Acad Sci D.*; 260:7033-7036

Levy JA, Hoffmann AD, Kramer SM, Landis JM, Shimabukuro JM, Oshiro LS (1984) Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science*; 225: 840-842

Lewis DA (2000) Retroviruses and the pathogenesis of schizophrenia. *Proc Natl Acad Sci USA*; 98:4293-4294.

Löwer R, Boller K, Hasenmaier B, Korbmacher C, Müller-Lantzsch N, Löwer J, Kurth R (1993) Identification of endogenous retroviruses with complex mRNA expression and particle formation. *Proc. Natl. Acad. Sci.*; 90: 4480-4484

Löwer R, Löwer J, Kurth R (1996) The viruses in all of us: Characteristics and biological significance of human endogenous retrovirus sequences. *Proc. Natl. Acad. Sci.*; 93: 5177-5184

Löwer R, Löwer J, Tondera-Koch C, Kurth R (1993) A general method of identification of transcribed retroviral sequences (R-U5 PCR) reveals the expression of the human endogenous retrovirus loci HERV-H and HERV-K in teratocarcinoma cells. *Virology*; 192,501–511.

Magin C, Löwer R, Löwer J (1999) cORF and RcRE, the Rev/Rex and RRE/RxRE homologues of the human endogenous retrovirus family HTDV/HERV-K. *J. Virol.*; 73 9496–9507.

Marchuck D, Drum M, Salino A, Collins FS (1991) Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucl. Acid Res.*, 19:1154

- Mariani-Costantini R, Horn TB, Callahan R (1989) Ancestry of a human endogenous retrovirus family. *J. Virol.*; 63:4982-4985
- Maas S (1999) Efficient and rapid procedure for blue-white screening of recombinant bacterial clones. *Biotechniques* 27:1126-1128
- Mayer J, Stuhr T, Reus K, Maldener E, Kitova M, Asmus F, Meese E (2005) Haplotype analysis of the human endogenous retrovirus locus HERV-K(HML-2.HOM) and its evolutionary implications. *J Mol Evol.*; 61:706-715
- Mayer J, Sauter M, Rácz A, Scherer D, Mueller-Lantzsch N, Meese E (1999) An almost-intact human endogenous retrovirus K on human chromosome 7. *Nat Genet.*; 21:257-258
- Medstrand P, Blomberg J (1993) Characterization of novel reverse transcriptase encoding human endogenous retroviral sequences similar to type A and type B retroviruses: differential transkription in normal human tissues. *J. Virol.*; 67:6778-6787
- Moyes DL, Martin A, Sawcer S, Temperton N, Worthington J, Griffiths DJ, Venables PJ. (2005) The distribution of the endogenous retroviruses HERV-K113 and HERV-K115 in health and disease. *Genomics*; 86:337-341
- Müllенbach R, Lagoda PJ, Welter C (1989) An efficient salt-chloroform extraction of DNA from blood and tissues. *Trends Genet.* 5:391
- Müller-Lantzsch N, Sauter M, Weiskircher A, Kramer K, Best B, Buck M, Grässer F (1993) Human endogenous retroviral element K10 (HERV-K10) encodes a full-length gag homologous 73-kDa protein and a functional protease. *AIDS Res. Hum. Retroviruses* 9:343-350
- Nelson PN, Carnegie PR, Martin J, Davari Ejtehadi H, Hooley P, Roden D, owland-Jones S, Warren P, Astley J, Murray PG (2003) Demystified...Human endogenous retroviruses. *J. Clin. Pathol: Mol. Pathol.*; 56:11-18

Ono M (1986) Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. *J. Virol.*; 58:937-944

Oroszlan S, Luftig RB (1990) Retroviral proteinases. *Curr. Top. Microbiol. Immunol.*; 157:153-185

Otowa T, Tochigi M, Rogers M, Umekage T, Kato N, Sasaki T (2006) Insertional polymorphism of endogenous retrovirus HERV-K115 in schizophrenia. *Neurosci Lett.*; 408:226-229

Perron H, Seigneurin JM (1999) Human retroviral sequences associated with extracellular particles in autoimmune diseases: epiphenomenon or possible role in aetiopathogenesis? *Microbes Infect.*; 1:309-322

Poiesz BJ, Poiesz MJ, Choi D (2003) The human T-cell lymphoma/leukemia viruses. *Cancer Invest*; 21:253

Ponferrada VG, Mauck BS, Wooley DP (2003) The envelope glycoprotein of human endogenous retrovirus HERV-W induces cellular resistance to spleen necrosis virus, *Arch Virol.*; 148:659-75

Rabbi MF, Saifuddin M, Gu DS, Kagnoff MF, Roebuck KA (1997) U5 Region of the Human Immunodeficiency Virus Type 1 Long Terminal Repeat Contains TRE-like cAMP-Responsive Elements that Bind both AP-1 and CREB/ATF Proteins. *Virology*; 233: 235-245

Rasmussen HB, Geny C, Deforges L, Perron H, Tourtelotte W, Heltberg A, Clausen J (1995) Expression of endogenous retroviruses in blood mononuclear cells and brain tissue from multiple sclerosis patients. *Mult Scler.*; 1:82-87

Reus K, Mayer J, Sauter M, Scherer D, Muller-Lantzsch N, Meese E (2001) Genomic organization of the human endogenous retrovirus HERV-K(HML-2.HOM) (ERV-K6) on chromosome 7. *Genomics*; 72:314-320

Rous P (1911) A sarcoma of the fowl transmissible by an agent separable from the tumor cells. *J. Exp. Med.*; 13:397-411

Rozen S, Skaletsky (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds.) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp.365-386

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*; 230:1350-1354

Sauter M, Schommer S, Kremmer E, Remberger K, Dolken G, Lemm I, Buck M, Best B, Neumann-Haefelin D, Müller-Lantzsch N (1995) Human endogenous retrovirus K10: expression of Gag protein and detection of antibodies in patients with seminomas. *J Virol.*; 69:414-421

Sauter M, Roemer K, Best B, Afting M, Schommer S, Seitz G, Hartmann M, Müller-Lantzsch N (1996) Specificity of antibodies directed against Env protein of human endogenous retroviruses in patients with germ cell tumors. *Cancer Res.*; 56:4362-4365

Schmelz HU, Albers P (2007) Hodentumore In: *Facharztwissen Urologie Differenzierte Diagnostik und Therapie* Schmelz HU, Sparwasser C, Weidner W (eds.) Springer-Verlag Berlin, pp.258-281

Shaknovich R, Yeyati PL, Ivins S, Melnick A, Lempert C, Waxman S, Zelent A, Licht JD (1998) The promyelocytic leukemia zinc finger protein affects myeloid cell growth, differentiation, and apoptosis. *Mol Cell Biol.* 18:5533-5545

Sigurdsson B (1954) Observations on three slow infections of sheep. Maedi, paratuberculosis, rida, a slow encephalitis of sheep with general remarks on infections which develop slowly and some of their special characteristics. *Br. Vet. J.*; 110: 255-270

Steele PE, Martin MA, Rabson AB, Bryan T, O'Brien SJ (1986) Amplification and chromosomal dispersion of human endogenous retroviral sequences. *J. Virol.*; 59: 545-550

Steinhuber S, Brack M, Hunsmann G, Schwelberger H, Dierich M, Vogetseder W (1995) Distribution of human endogenous retrovirus HERV-K genomes in humans and different primates. *Hum Genet* ; 96:188-192

Stuhr T. (2004) Humane endogene Retrovirus Familie HERV-K(HML-2): Polymorphismen eines Provirus im Menschen und Frequenz intakter homologer retroviraler Gene in niederen Altweltaffen. Dissertation; Institut für Humangenetik der Universität des Saarlandes

Takeuchi Y, Vile RG, Simpson G, O'Hara B, Collins MK, Weiss RA (1992) Feline leukemia virus subgroup B uses the same cell surface receptor as gibbon ape leukemia virus. *J. Virol.*;66: 1219-1222

Temin HM, Mizutani S (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* ; 226:1211-1213

Ting CN, Rosenberg MP, Snow CM, Samuelson LC, Meisler MH (1992) Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev.* 6:1457-1465

Todaro GJ, Scherr CJ, Benveniste RR, Lieber MM (1974) Type C viruses of baboons: Isolation from normal cell cultures. *Cell*; 2:55-61

Turner G, Barbulescu M, Mei S, Jensen-Seaman MI, Kidd KK, Lenz J (2001) Insertional polymorphisms of full-length endogenous retroviruses in humans. *Curr Biol*, 11; 1531-1535

Ullmann A, Perrin D, Jacob F, Monod J (1965) Identification, by in vitro complementation and purification, of a peptide fraction of Escherichia coli beta-galactosidase. *J Mol Biol.* 12: 918-923

Vallée H, Carré H (1904) Sur la nature infectieuse de l'anémie du cheval. *C.R. Hebd. Seances Acad. Sci. Ser. D Sci. Nat.*;139: 331-333

Van Nie R, Verstraeten AA, DeMoes J (1977) Genetic transmission of mammary tumor virus by Gr mice. *Int. J. Cancer*; 19: 383-390

Venables PJ (2004) Sjogren's syndrome. *Best Pract Res Clin Rheumatol.*; 18:313-29

Wang-Johanning F, Frost AR, Jian B, Azerou R, Lu DW, Chen DT, Johanning GL (2003) Detecting the expression of human endogenous retrovirus E envelope transcripts in human prostate adenocarcinoma, *Cancer*; 98:187-197

Wang-Johanning F, Frost AR, Jian B, Epp L, Lu DW, Johanning GL (2003) Quantitation of HERV-K env gene expression and splicing in human breast cancer. *Oncogene*; 22:1528-1535

Wang-Johanning F, Frost AR, Johanning GL, Khazaeli MB, LoBuglio AF, Shaw DR, Strong TV (2001) Expression of human endogenous retrovirus k envelope transcripts in human breast cancer, *Clin Cancer Res.* 7:1553-1560

Wang-Johanning F, Liu J, Rycak K, Huang M, Tsai K, Rosen DG, Chen DT, Lu DW, Barnhart KF, Johanning GL (2007) Expression of multiple human endogenous retrovirus surface envelope proteins in ovarian cancer. *Int J Cancer*; 120:81-90

Wilkinson AD, Mager DL, Leong JAC (1994) Endogenous human retroviruses. In Levy J.A., ed. *The Retroviridae*, Vol. 3. New York: Plenum Press pp.465-535

Yolken RH, Karlsson H, Yee F, Johnston-Wilson NL, Torrey EF (2000) Endogenous retroviruses and schizophrenia. *Brain Res Brain Res Rev.*; 31:193-199

## 7. Anhang

### 7.1 Die Sequenz der duplizierte Region

```

TS24      AAAAGCCATCGATTGTACAGTAATAAGTTCAATACAAAGCAGAGAATACATTTCTTTTA
P68      AAAAGCCATCGATTGTACAGTAATAAGTTCAATACAAAGCAGAGAATATACTTGCTTTCA
          ***** * * * * *

TS24      GCATTTTGTAGGCTTTTAGTTTCTAGTAGTCATCTTGTTAAAATGATTGTTTGTTC
P68      GCATTTTAAAGGTTTCTAGTTTCTAGTAGTCATCTTGTTGAAATGATTGCCTGTTCA
          ***** * * * * *

TS24      GTATTGCTTTTCTTCGGAACAGGTACAAATTCATAC--ACACAACTCACT---C
P68      ATATTG--TCTTCTCTGAAAATAGTTACAAATTCATACTTACACAACTCACTTACTC
          .***** * * * * *

TS24      TATAATTTTCTTACACTTAAGGTTTATCTTTTGACTAACATATATTTAACTATATGTAAA
P68      TATAATTTTCTTACACCTAAGGTTTATCTTTTGACTAACATATATTTGAAGTCTATGTAAA
          ***** * * * * *

TS24      TCAAAACCTAAACGTCTGCATGTGTAA
P68      TCAAAACCTAAACGTCTGCATGTGT--
          *****

```

TS 24 repräsentiert die Sequenz der duplizierten Region. P68 spiegelt die Sequenz der Präintegrationsstelle wider. Die Schnittstellen für *Afl*III in der duplizierten Region und *Ssp*I in der flankierenden Sequenz sind markiert.

### 7.2 Das sequenzierte *env*-Gen

```

K113_env  --ATGAACCCATCGGAGATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACACCGCAATC
P74      -TATGAACCCATCGGAGATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACACCGCAATC
P2       -TATGAACCCATCGGAGATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACACCGCAATC
P69      -TATGAACCCATCGGAGATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACACCGCAATC
P24      -TATGAACCCATCGGAGATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACACCGCAATC
P23      --ATGAACCCATCGGAGATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACACCGCAATC
P76      TTATGAACCCATCGGAGATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACACCGCAATC
P28      --ATGAACCCATCGGAGATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACACCGCAATC
          *****

K113_env  GAGCACCGTTGACTCACAAGATGAACAAAATGGTGACGTCAGAAGAACAGATGAAGTTGC
P74      GAGCACCGTTGACTCACAAGATGAACAAAATGGTGACGTCAGAAGAACAGATGAAGTTGC
P2       GAGCACCGTTGACTCACAAGATGAACAAAATGGTGACGTCAGAAGAACAGATGAAGTTGC
P69      GAGCACCGTTGACTCACAAGATGAACAAAATGGTGACGTCAGAAGAACAGATGAAGTTGC
P24      GAGCACCGTTGACTCACAAGATGAACAAAATGGTGACGTCAGAAGAACAGATGAAGTTGC
P23      GAGCACCGTTGACTCACAAGATGAACAAAATGGTGACGTCAGAAGAACAGATGAAGTTGC
P76      GAGCACCGTTGACTCACAAGATGAACAAAATGGTGACGTCAGAAGAACAGATGAAGTTGC
P28      GAGCACCGTTGACTCACAAGATGAACAAAATGGTGACGTCAGAAGAACAGATGAAGTTGC
          *****

K113_env  CATCCACCAAGAAGGCAGAGCCGCCGACTTGGGCACAACCTAAAGAAGCTGACGCAGTTAG
P74      CATCCACCAAGAAGGCAGAGCCGCCGACTTGGGCACAACCTAAAGAAGCTGACGCAGTTAG
P2       CATCCACCAAGAAGGCAGAGCCGCCGACTTGGGCACAACCTAAAGAAGCTGACGCAGTTAG
P69      CATCCACCAAGAAGGCAGAGCCGCCGACTTGGGCACAACCTAAAGAAGCTGACGCAGTTAG
P24      CATCCACCAAGAAGGCAGAGCCGCCGACTTGGGCACAACCTAAAGAAGCTGACGCAGTTAG

```

P23 CATCCACCAAGAAGGCAGAGCCGCCGACTTGGGCACAACCTAAAGAAGCTGACGCAGTTAG  
P76 CATCCACCAAGAAGGCAGAGCCGCCGACTTGGGCACAACCTAAAGAAGCTGACGCAGTTAG  
P28 CATCCACCAAGAAGGCAGAGCCGCCGACTTGGGCACAACCTAAAGAAGCTGACGCAGTTAG  
\*\*\*\*\*

K113\_env  
P74 CTACAAAATATCTAGAGAACACAAAGGTGACACAAACCCAGAGAGTATGCTGCTTGCAG  
P2 CTACAAAATATCTAGAGAACACAAAGGTGACACAAACCCAGAGAGTATGCTGCTTGCAG  
P69 CTACAAAATATCTAGAGAACACAAAGGTGACACAAACCCAGAGAGTATGCTGCTTGCAG  
P24 CTACAAAATATCTAGAGAACACAAAGGTGACACAAACCCAGAGAGTATGCTGCTTGCAG  
P23 CTACAAAATATCTAGAGAACACAAAGGTGACACAAACCCAGAGAGTATGCTGCTTGCAG  
P76 CTACAAAATATCTAGAGAACACAAAGGTGACACAAACCCAGAGAGTATGCTGCTTGCAG  
P28 CTACAAAATATCTAGAGAACACAAAGGTGACACAAACCCAGAGAGTATGCTGCTTGCAG  
\*\*\*\*\*

K113\_env  
P74 CCTTGATGATTGTATCAATGGTGGTAAGTCTCCCTATGCCTGCAGGAGCAGCTGCAGCTA  
P2 CCTTGATGATTGTATCAATGGTGGTAAGTCTCCCTATGCCTGCAGGAGCAGCTGCAGCTA  
P69 CCTTGATGATTGTATCAATGGTGGTAAGTCTCCCTATGCCTGCAGGAGCAGCTGCAGCTA  
P24 CCTTGATGATTGTATCAATGGTGGTAAGTCTCCCTATGCCTGCAGGAGCAGCTGCAGCTA  
P23 CCTTGATGATTGTATCAATGGTGGTAAGTCTCCCTATGCCTGCAGGAGCAGCTGCAGCTA  
P76 CCTTGATGATTGTATCAATGGTGGTAAGTCTCCCTATGCCTGCAGGAGCAGCTGCAGCTA  
P28 CCTTGATGATTGTATCAATGGTGGTAAGTCTCCCTATGCCTGCAGGAGCAGCTGCAGCTA  
\*\*\*\*\*

K113\_env  
P74 ACTATACCTACTGGGCCTATGTGCCTTTCCCGCCCTTAATTCGGGCAGTCACATGGATGG  
P2 ACTATACCTACTGGGCCTATGTGCCTTTCCCGCCCTTAATTCGGGCAGTCACATGGATGG  
P69 ACTATACCTACTGGGCCTATGTGCCTTTCCCGCCCTTAATTCGGGCAGTCACATGGATGG  
P24 ACTATACCTACTGGGCCTATGTGCCTTTCCCGCCCTTAATTCGGGCAGTCACATGGATGG  
P23 ACTATACCTACTGGGCCTATGTGCCTTTCCCGCCCTTAATTCGGGCAGTCACATGGATGG  
P76 ACTATACCTACTGGGCCTATGTGCCTTTCCCGCCCTTAATTCGGGCAGTCACATGGATGG  
P28 ACTATACCTACTGGGCCTATGTGCCTTTCCCGCCCTTAATTCGGGCAGTCACATGGATGG  
\*\*\*\*\*

K113\_env  
P74 ATAATCCTATAGAAATATATGTTAATGATAGTGTATGGGTACCTGGCCCCACAGATGATT  
P2 ATAATCCTATAGAAATATATGTTAATGATAGTGTATGGGTACCTGGCCCCACAGATGATT  
P69 ATAATCCTATAGAAATATATGTTAATGATAGTGTATGGGTACCTGGCCCCACAGATGATT  
P24 ATAATCCTATAGAAATATATGTTAATGATAGTGTATGGGTACCTGGCCCCACAGATGATT  
P23 ATAATCCTATAGAAATATATGTTAATGATAGTGTATGGGTACCTGGCCCCACAGATGATT  
P76 ATAATCCTATAGAAATATATGTTAATGATAGTGTATGGGTACCTGGCCCCACAGATGATT  
P28 ATAATCCTATAGAAATATATGTTAATGATAGTGTATGGGTACCTGGCCCCACAGATGATT  
\*\*\*\*\*

K113\_env  
P74 GCTGCCCTGCCAAACCTGAGGAAGAAGGGATGATGATAAATATTTCCATTGGGTATCGTT  
P2 GCTGCCCTGCCAAACCTGAGGAAGAAGGGATGATGATAAATATTTCCATTGGGTATCGTT  
P69 GCTGCCCTGCCAAACCTGAGGAAGAAGGGATGATGATAAATATTTCCATTGGGTATCGTT  
P24 GCTGCCCTGCCAAACCTGAGGAAGAAGGGATGATGATAAATATTTCCATTGGGTATCGTT  
P23 GCTGCCCTGCCAAACCTGAGGAAGAAGGGATGATGATAAATATTTCCATTGGGTATCGTT  
P76 GCTGCCCTGCCAAACCTGAGGAAGAAGGGATGATGATAAATATTTCCATTGGGTATCGTT  
P28 GCTGCCCTGCCAAACCTGAGGAAGAAGGGATGATGATAAATATTTCCATTGGGTATCGTT  
\*\*\*\*\*

K113\_env  
P74 ATCCTCCTATTTGCCTAGGGAGAGCACCAGGATGTTTAAATGCCTGCAGTCCAAAATTGGT  
P2 ATCCTCCTATTTGCCTAGGGAGAGCACCAGGATGTTTAAATGCCTGCAGTCCAAAATTGGT  
P69 ATCCTCCTATTTGCCTAGGGAGAGCACCAGGATGTTTAAATGCCTGCAGTCCAAAATTGGT  
P24 ATCCTCCTATTTGCCTAGGGAGAGCACCAGGATGTTTAAATGCCTGCAGTCCAAAATTGGT  
P23 ATCCTCCTATTTGCCTAGGGAGAGCACCAGGATGTTTAAATGCCTGCAGTCCAAAATTGGT  
P76 ATCCTCCTATTTGCCTAGGGAGAGCACCAGGATGTTTAAATGCCTGCAGTCCAAAATTGGT  
P28 ATCCTCCTATTTGCCTAGGGAGAGCACCAGGATGTTTAAATGCCTGCAGTCCAAAATTGGT  
\*\*\*\*\*



```

K113_env      TGGTAGAAGTACCTACTGTCTAGTCCCATCAGTAGATTCACTTATCACATGGTAAGCGGGA
P74           TGGTAGAAGTACCTACTGTCTAGTCCCATCAGTAGATTCACTTATCACATGGTAAGCGGGA
P2            TGGTAGAAGTACCTACTGTCTAGTCCCATCAGTAGATTCACTTATCACATGGTAAGCGGGA
P69           TGGTAGAAGTACCTACTGTCTAGTCCCATCAGTAGATTCACTTATCACATGGTAAGCGGGA
P24           TGGTAGAAGTACCTACTGTCTAGTCCCATCAGTAGATTCACTTATCACATGGTAAGCGGGA
P23           TGGTAGAAGTACCTACTGTCTAGTCCCATCAGTAGATTCACTTATCACATGGTAAGCGGGA
P76           TGGTAGAAGTACCTACTGTCTAGTCCCATCAGTAGATTCACTTATCACATGGTAAGCGGGA
P28           TGGTAGAAGTACCTACTGTCTAGTCCCATCAGTAGATTCACTTATCACATGGTAAGCGGGA
*****

K113_env      TGTCACCTCAGGCCACGGGTAAATTATTTACAAGACTTTTCTTATCAAAGATCATTTAAAT
P74           TGTCACCTCAGGCCACGGGTAAATTATTTACAAGACTTTTCTTATCAAAGATCATTTAAAT
P2            TGTCACCTCAGGCCACGGGTAAATTATTTACAAGACTTTTCTTATCAAAGATCATTTAAAT
P69           TGTCACCTCAGGCCACGGGTAAATTATTTACAAGACTTTTCTTATCAAAGATCATTTAAAT
P24           TGTCACCTCAGGCCACGGGTAAATTATTTACAAGACTTTTCTTATCAAAGATCATTTAAAT
P23           TGTCACCTCAGGCCACGGGTAAATTATTTACAAGACTTTTCTTATCAAAGATCATTTAAAT
P76           TGTCACCTCAGGCCACGGGTAAATTATTTACAAGACTTTTCTTATCAAAGATCATTTAAAT
P28           TGTCACCTCAGGCCACGGGTAAATTATTTACAAGACTTTTCTTATNNNAGATCATTTAAAT
*****

K113_env      TTAGACCTAAAGGGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAATACAGAAG
P74           TTAGACCTAAAGGGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAATACAGAAG
P2            TTAGACCTAAAGGGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAATACAGAAG
P69           TTAGACCTAAAGGGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAATACAGAAG
P24           TTAGACCTAAAGGGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAATACAGAAG
P23           TTAGACCTAAAGGGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAATACAGAAG
P76           TTAGACCTAAAGGGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAATACAGAAG
P28           TTAGACCTAAAGGGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAATACAGAAG
*****

K113_env      TTTTAGTTTGGGAAGAATGTGTGGCCAATAGTGCGGTGATATTACAAAACAATGAATTTCG
P74           TTTTAGTTTGGGAAGAATGTGTGGCCAATAGTGCGGTGATATTACAAAACAATGAATTTCG
P2            TTTTAGTTTGGGAAGAATGTGTGGCCAATAGTGCGGTGATATTACAAAACAATGAATTTCG
P69           TTTTAGTTTGGGAAGAATGTGTGGCCAATAGTGCGGTGATATTACAAAACAATGAATTTCG
P24           TTTTAGTTTGGGAAGAATGTGTGGCCAATAGTGCGGTGATATTACAAAACAATGAATTTCG
P23           TTTTAGTTTGGGAAGAATGTGTGGCCAATAGTGCGGTGATATTACAAAACAATGAATTTCG
P76           TTTTAGTTTGGGAAGAATGTGTGGCCAATAGTGCGGTGATATTACAAAACAATGAATTTCG
P28           TTTTAGTTTGGGAAGAATGTGTGGCCAATAGTGCGGTGATATTACAAAACAATGAATTTCG
*****

K113_env      GAACTCTTATAGATTGGGCACCTCGAGGTCAATTCTACCACAATTGCTCAGGACAAACTC
P74           GAACTCTTATAGATTGGGCACCTCAAGGTCAATTCTACCACAATTGCTCAGGACAAACTC
P2            GAACTCTTATAGATTGGGCACCTCAAGGTCAATTCTACCACAATTGCTCAGGACAAACCC
P69           GAACTCTTATAGATTGGGCACCTCAAGGTCAATTCTACCACAATTGCTCAGGACAAACTC
P24           GAACTCTTATAGATTGGGCACCTCAAGGTCAATTCTACCACAATTGCTCAGGACAAACTC
P23           GAACTCTTATAGATTGGGCACCTCAAGGTCAATTCTACCACAATTGCTCAGGACAAACTC
P76           GAACTCTTATAGATTGGGCACCTCAAGGTCAATTCTACCACAATTGCTCAGGACAAACTC
P28           GAACTCTTATAGATTGGGCACCTCAAGGTCAATTCTACCACAATTGCTCAGGACAAACTC
*****

K113_env      AGTCGTGTCCAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACTTAACAGAAAGTTTAG
P74           AGTCGTGTCCAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACTTAACAGAAAGTTTAG
P2            AGTCGTGTCCAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACTTAACAGAAAGTTTAG
P69           AGTCGTGTCCAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACTTAACAGAAAGTTTAG
P24           AGTCGTGTCCAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACTTAACAGAAAGTTTAG
P23           AGTCGTGTCCAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACTTAACAGAAAGTTTAG
P76           AGTCGTGTCCAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACTTAACAGAAAGTTTAG
P28           AGTCGTGTCCAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACTTAACAGAAAGTTTAG
*****

K113_env      ACAACATAAGCATAAAAAATTGCAGTCTTCTACCCTTGGGAATGGGGAGAAAAAGGAA
P74           ACAACATAAGCATAAAAAATTGCAGTCTTCTACCCTTGGGAATGGGGAGAAAAAGGAA
P2            ACAACATAAGCATAAAAAATTGCAGTCTTCTACCCTTGGGAATGGGGAGAAAAAGGAA
P69           ACAACATAAGCATAAAAAATTGCAGTCTTCTACCCTTGGGAATGGGGAGAAAAAGGAA
P24           ACAACATAAGCATAAAAAATTGCAGTCTTCTACCCTTGGGAATGGGGAGAAAAAGGAA
P23           ACAACATAAGCATAAAAAATTGCAGTCTTCTACCCTTGGGAATGGGGAGAAAAAGGAA
P76           ACAACATAAGCATAAAAAATTGCAGTCTTCTACCCTTGGGAATGGGGAGAAAAAGGAA
P28           ACAACATAAGCATAAAAAATTGCAGTCTTCTACCCTTGGGAATGGGGAGAAAAAGGAA
*****

```

```

K113_env      TCTCTACCGCAAGACCAAAAAATAATAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTAT
P74           TCTCTACCGCAAGACCAAAAAATAATAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTAT
P2            TCTCTACCGCAAGACCAAAAAATAATAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTAT
P69           TCTCTACCGCAAGACCAAAAAATAATAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTAT
P24           TCTCTACCGCAAGACCAAAAAATAATAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTAT
P23           TCTCTACCGCAAGACCAAAAAATAATAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTAT
P76           TCTCTACCGCAAGACCAAAAAATAATAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTAT
P28           TCTCTATCGCAAGACCAAAAAATAATAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTAT
*****  **  *****

K113_env      GGAGGCTTACTGTGGCCTCACACCACATTAGAATTTGGTCTGGAAATCAAACTTTAGAAA
P74           GGAGGTTTACTGTGGCCTCACACCACATTAGAATTTGGTCTGGAAATCAAACTTTAGAAA
P2            GGAGGCTTACTGTGGCCTCACACCACATTAGAATTTGGTCTGGAAATCAAACTTTAGAAA
P69           GGAGGCTTACTGTGGCCTCACACCACATTAGAATTTGGTCTGGAAATCAAACTTTAGAAA
P24           GGAGGCTTACTGTGGCCTCACACCACATTAGAATTTGGTCTGGAAATCAAACTTTAGAAA
P23           GGAGGCTTACTGTGGCCTCACACCACATTAGAATTTGGTCTGGAAATCAAACTTTAGAAA
P76           GGAGGCTTACTGTGGCCTCACACCACATTAGAATTTGGTCTGGAAATCAAACTTTAGAAA
P28           GGAGGCTTACTGTGGCCTCACACCACATTAGAATTTGGTCTGGAAATCAAACTTTAGAAA
*****  *****

K113_env      CAAGAGATCGTAAGCCATTTTATACTATCGACCTAAATTCAGTCTAACAGTTCCTTTAC
P74           CAAGAGATCGTAAGCCATTTTATACTATCGACCTAAATTCAGTCTAACAGTTCCTTTAC
P2            CAAGAGATCGTAAGCCATTTTATACTATCGACCTAAATTCAGTCTAACAGTTCCTTTAC
P69           CAAGAGATCGTAAGCCATTTTATACTATCGACCTAAATTCAGTCTAACAGTTCCTTTAC
P24           CAAGAGATCGTAAGCCATTTTATACTATCGACCTAAATTCAGTCTAACAGTTCCTTTAC
P23           CAAGAGATCGTAAGCCATTTTATACTATCGACCTAAATTCAGTCTAACAGTTCCTTTAC
P76           CAAGAGATCGTAAGCCATTTTATACTATCGACCTAAATTCAGTCTAACAGTTCCTTTAC
P28           CAAGAGATCGTAAGCCATTTTATACTATCGACCTAAATTCAGTCTAACAGTTCCTTTAC
*****

K113_env      AAAGTTGCGTAAAGCCCCCTTATATGCTAGTTGTAGGAAATATAGTTATTAACCAGACT
P74           AAAGTTGCGTAAAGCCCCCTTATATGCTAGTTGTAGGAAATATAGTTATTAACCAGACT
P2            AAAGTTGCGTAAAGCCCCCTTATATGCTAGTTGTAGGAAATATAGTTATTAACCAGACT
P69           AAAGTTGCGTAAAGCCCCCTTATATGCTAGTTGTAGGAAATATAGTTATTAACCAGACT
P24           AAAGTTGCGTAAAGCCCCCTTATATGCTAGTTGTAGGAAATATAGTTATTAACCAGACT
P23           AAAGTTGCGTAAAGCCCCCTTATATGCTAGTTGTAGGAAATATAGTTATTAACCAGACT
P76           AAGGTTGCGTAAAGCCCCCTTATATGCTAGTTGTAGGAAATATAGTTATTAACCAGACT
P28           AAAGTTGCGTAAAGCCCCCTTATATGCTAGTTGTAGGAAATATAGTTATTAACCAGACT
** , *****

K113_env      CCCAGACTATAACCTGTGAAAATTGTAGATTGCTTACTTGCATTGATTCAACTTTTAATT
P74           CCCAGACTATAACCTGTGAAAATTGTAGATTGCTTACTTGCATTGATTCAACTTTTAATT
P2            CCCAGACTATAACCTGTGAAAATTGTAGATTGCTTACTTGCATTGATTCAACTTTTAATT
P69           CCCAGACTATAACCTGTGAAAATTGTAGATTGCTTACTTGCATTGATTCAACTTTTAATT
P24           CCCAGACTATAACCTGTGAAAATTGTAGATTGCTTACTTGCATTGATTCAACTTTTAATT
P23           CCCAGACTATAACCTGTGAAAATTGTAGATTGCTTACTTGCATTGATTCAACTTTTAATT
P76           CCCAGACTATAACCTGTGAAAATTGTAGATTGCTTACTTGCATTGATTCAACTTTTAATT
P28           CCCAGACTATAACCTGTGAAAATTGTAGATTGCTTACTTGCATTGATTCAACTTTTAATT
*****

K113_env      GGCAACACCGTATTCTGCTGGTGAGAGCAAGAGAGGGCGTGTGGATCCCTGTGTCCATGG
P74           GGCAACACCGTATTCTGCTGGTGAGAGCAAGAGAGGGCGTGTGGATCCCTGTGTCCATGG
P2            GGCAACACCGTATTCTGCTGGTGAGAGCAAGAGAGGGCGTGTGGATCCCTGTGTCCATGG
P69           GGCAACACCGTATTCTGCTGGTGAGAGCAAGAGAGGGCGTGTGGATCCCTGTGTCCATGG
P24           GGCAACACCGTATTCTGCTGGTGAGAGCAAGAGAGGGCGTGTGGATCCCTGTGTCCATGG
P23           GGCAGCACCGTATTCTGCTGGTGAGAGCAAGAGAGGGCGTGTGGATCCCTGTGTCCATGG
P76           GGCAACACCGTATTCTGCTGGTGAGAGCAAGAGAGGGCGTGTGGATCCCTGTGTCCATGG
P28           GGCAACACCGTATTCTGCTGGTGAGAGCAAGAGAGGGCGTGTGGATCCCTGTGTCCATGG
**** , *****

K113_env      ACCGACCGTGGGAGGCCTCACCATCCGTCCATATTTTGACTGAAGTATTAAGGTGTTT
P74           ACCGACCGTGGGAGGCCTCACCATCCGTCCATATTTTGACTGAAGTATTAAGGTGTTT
P2            ACCGACCGTGGGAGGCCTCACCATCCGTCCATATTTTGACTGAAGTATTAAGGTGTTT
P69           ACCGACCGTGGGAGGCCTCACCATCCGTCCATATTTTGACTGAAGTATTAAGGTGTTT
P24           ACCGACCGTGGGAGGCCTCACCATCCGTCCATATTTTGACTGAAGTATTAAGGTGTTT
P23           ACCGACCGTGGGAGGCCTCACCATCCGTCCATATTTTGACTGAAGTATTAAGGTGTTT
P76           ACCGACCGTGGGAGGCCTCACCATCCGTCCATATTTTGACTGAAGTATTAAGGTGTTT
P28           ACCGACCGTGGGAGGCCTCACCATCCGTCCATATTTTGACTGAAGTATTAAGGTGTTT
*****  *****

```

```

K113_env      TAAATAGATCCAAAAGATTCAATTTTACTTTAATTGCAGTGATTATGGGATTAATTGCAG
P74           TAAATAGATCCAAAAGATTCAATTTTACTTTAATTGCAGTGATTATGGGATTAATTGCAG
P2            TAAATAGATCCNAAAGATTCAATTTTACTTTAATTGCAGTGATTATGGGATTAATTGCAG
P69           TAAATAGATCCAAAAGATTCAATTTTACTTTAATTGCAGTGATTATGGGATTAATTGCAG
P24           TAAATAGATCCAAAAGATTCAATTTTACTTTAATTGCAGTGATTATGGGATTAATTGCAG
P23           TAAATAGATCCAAAAGATTCAATTTTACTTTAATTGCAGTGATTATGGGATTAATTGCAG
P76           TAAATAGATCCAAAAGATTCAATTTTACTTTAATTGCAGTGATTATGGGATTAATTGCAG
P28           TAAATAGATCCAAAAGATTCAATTTTACTTTAATTGCAGTGATTATGGGATTAATTGCAG
*****

K113_env      TCACAGCTACGGCTGCTGTAGCAGGAGTTGCATTGCACTCTTCTGTTTCAGTCAGTAAACT
P74           TCACAGCTACGGCTGCTGTAGCAGGACTTGCATTGCACTCTTCTTTTCAGTCAGTAAACT
P2            TCACAGCTACGGCTGCTGTAGCAGGACTTGCATTGCACTCTTCTTTTCAGTCAGTAAACT
P69           TCACAGCTACGGCTGCTGTAGCAGGACTTGCATTGCACTCTTCTTTTCAGTCAGTAAACT
P24           TCACAGCTACGGCTGCTGTAGCAGGACTTTCATTGCACTCTTCTTTTCAGTCAGTAAACT
P23           TCACAGCTACGGCTGCTGTAGCAGGACTTGCATTGCACTCTTCTTTTCAGTCAGTAAACT
P76           TCACAGCTACGGCTGCTGTAGCAGGACTTGCATTGCACTCTTCTTTTCAGTCAGTAAACT
P28           TCACAGCTACGGCTGCTGTAGCAGGACTTGCATTGCACTCTTCTTTTCAGTCAGTAAACT
*****

K113_env      TTGTTAATGATTGGCAAAATAATTCTACAAGATTGTGGAATTCACAATCTAGTATTGATC
P74           TTGTTAATGATTGGCAAAATAATTCTACAAGATTGTGGAATTCACAATCTAGTATTGATC
P2            TTGTTAATGATTGGCAAAATAATTCTACAAGATTGTGGAATTCACAATCTAGTATTGATC
P69           TTGTTAATGATTGGCAAAATAATTCTACAAGATTGTGGAATTCACAATCTAGTATTGATC
P24           TTGTTAATGATTGGCAAAATAATTCTACAAGATTGTGGAATTCACAATCTAGTATTGATC
P23           TTGTTAATGATTGGCAAAATAATTCTACAAGATTGTGGAATTCACAATCTAGTATTGATC
P76           TTGTTAATGATTGGCAAAATAATTCTACAAGATTGTGGAATTCACAATCTAGTATTGATC
P28           TTGTTAATGATTGGCAAAATAATTCTACAAGATTGTGGAATTCACAATCTAGTATTGATC
*****

K113_env      AAAAAATTGGCAAAATCAAATTAATGATCTTAGACAAACTGTCATTTGGATGGGAGACAGGC
P74           AAAAAATTGGCAAAATCAAATTAATGATCTTAGACAAACTGTCATTTGGATGGGAGACAGGC
P2            AAAAAATTGGCAAAATCAAATTAATGATCTTAGACAAACTGTCATTTGGATGGGAGACAGGC
P69           AAAAAATTGGCAAAATCAAATTAATGATCTTAGACAAACTGTCATTTGGATGGGAGACAGGC
P24           AAAAAATTGGCAAAATCAAATTAATGATCTTAGACAAACTGTCATTTGGATGGGAGACAGGC
P23           AAAAAATTGGCAAAATCAAATTAATGATCTTAGACAAACTGTCATTTGGATGGGAGACAGGC
P76           AAAAAATTGGCAAAATCAAATTAATGATCTTAGACAAACTGTCATTTGGATGGGAGACAGGC
P28           AAAAAATTGGCAAAATCAAATTAATGATCTTAGACAAACTGTCATTTGGATGGGAGACAGGC
*****

K113_env      TCATGAGCTTAGAACATCGTTTCCAGTTACAATGTGACTGGAATACGTCAGATTTTGTGA
P74           ACATGAGCTTAGAACATCGTTTCCAGTTACAATGTGACTGGAATACGTCAGATTTTGTGA
P2            ACATGAGCTTAGAACATCGTTTCCAGTTACAATGTGACTGGAATACGTCAGATTTTGTGA
P69           ACATGAGCTTAGAACATCGTTTCCAGTTACAATGTGACTGGAATACGTCAGATTTTGTGA
P24           ACATGAGCTTAGAACATCGTTTCCAGTTACAATGTGACTGGAATACGTCAGATTTTGTGA
P23           ACATGAGCTTAGAACATCGTTTCCAGTTACAATGTGACTGGAATACGTCAGATTTTGTGA
P76           ACATGAGCTTAGAACATCGTTTCCAGTTACAATGTGACTGGAATACGTCAGATTTTGTGA
P28           ACATGAGCTTAGAACATCGTTTCCAGTTACAATGTGACTGGAATACGTCAGATTTTGTGA
: *****

K113_env      TTACACCCCAAATTTATAATGAGTCTGAGCATCACTGGGACATGGTTAGATGCCATCTAC
P74           TTACACCCCAAATTTATAATGAGTCTGAGCATCACTGGGACATGGTTAGACGCCATCTAC
P2            TTACACCCCAAATTTATAATGAGTCTGAGCATCACTGGGACATGGTTAGACGCCATCTAC
P69           TTACACCCCAAATTTATAATGAGTCTGAGCATCACTGGGACATGGTTAGACGCCATCTAC
P24           TTACACCCCAAATTTATAATGAGTCTGAGCATCACTGGGACATGGTTAGACGCCATCTAC
P23           TTACACCCCAAATTTATAATGAGTCTGAGCATCACTGGGACATGGTTAGACGCCATCTAC
P76           TTACACCCCAAATTTATAATGAGTCTGAGCATCACTGGGACATGGTTAGACGCCATCTAC
P28           TTACACCCCAAATTTATAATGAGTCTGAGCATCACTGGGACATGGTTAGACGCCATCTAC
*****

K113_env      AGGGAAGAGAAGATAATCTCACTTTAGACATTTCCAAATTAAAAGAACAATTTTGAAG
P74           AGGGAAGAGAAGATAATCTCACTTTAGACATTTCCAAATTAAAAGAACAATTTTGAAG
P2            AGGGAAGAGAAGATAATCTCACTTTAGACATTTCCAAATTAAAAGAACAATTTTGAAG
P69           AGGGAAGAGAAGATAATCTCACTTTAGACATTTCCAAATTAAAAGAACAATTTTGAAG
P24           AGGGAAGAGAAGATAATCTCACTTTAGACATTTCCAAATTAAAAGAACAATTTTGAAG
P23           AGGGAAGAGAAGATAATCTCACTTTAGACATTTCCAAATTAAAAGAACAATTTTGAAG
P76           AGGGAAGAGAAGATAATCTCACTTTAGGCAATTTCCAAATTAAAAGAACAATTTTGAAG
P28           AGGGAAGAGAAGATAATCTCACTTTAGACATCTCCAAATTAAAAGAACAATTTTGAAG
*****

```

```

K113_env      CATCAAAAGCCCATTTAAATTTGGTGCCAGGAAC TGAGGCAATTGCAGGAGTTGCTGATG
P74           CATCAAAAGCCCATTTAAATTTGGTGCCAGGAAC TGAGGCAATTGCAGGAGTTGCTGATG
P2            CATCAAAAGCCCATTTAAATTTGGTGCCAGGAAC TGAGGCAATTGCAGGAGTTGCTGATG
P69           CATCAAAAGCCCATTTAAATTTGGTGCCAGGAAC TGAGGCAATTGCAGGAGTTGCTGATG
P24           CATCAAAAGCCCATTTAAATTTGGTGCCAGGAAC TGAGGCAATTGCAGGAGTTGCTGATG
P23           CATCAAAAGCCCATTTAAATTTGGTGCCAGGAAC TGAGGCAATTGCAGGAGTTGCTGATG
P76           CATCAAAAGCCCATTTAAATTTGGTGCCAGGAAC TGAGGCAATTGCAGGAGTTGCTGATG
P28           CATCAAAAGCCCATTTAAATTTGGTGCCAGGAAC TGAGGCAATTGCAGGAGTTGCTGATG
              *****

K113_env      GCCTCGCAAATCTTAACACTGTCAC TTGGGTTAAGACCATTGGAAGTACTACAATTATAA
P74           GCCTCGCAAATCTTAACACTGTCAC TTGGGTTAAGACCATTGGAAGTACTACAATTATAA
P2            GCCTCGCAAATCTTAACACTGTCAC TTGGGTTAAGACCATTGGAAGTACTACAATTATAA
P69           GCCTCGCAAATCTTAACACTGTCAC TTGGGTTAAGACCATTGGAAGTACTACAATTATAA
P24           GCCTCGCAAATCTTAACACTGTCAC TTGGGTTAAGACCATTGGAAGTACTACAATTATAA
P23           GCCTCGCAAATCTTAACACTGTCAC TTGGGTTAAGACCATTGGAAGTACTACAATTATAA
P76           GCCTCGCAAATCTTAACACTGTCAC TTGGGTTAAGACCATTGGAAGTACTACAATTATAA
P28           GCCTCGCAAATCTTAACACTGTCAC TTGGGTTAAGACCATTGGAAGTACTACAATTATAA
              *****

K113_env      ATCTCATATTAATCCTTGTGTGCCTGTTTGTCTGTTGTTAGTCTACAGGTGTACCCAAC
P74           ATCTCATATTAATCCTTGTGTGCCTGTTTGTCTGTTGTTAGTCTACAGGTGTACCCAAC
P2            ATCTCATATTAATCCTTGTGTGCCTGTTTGTCTGTTGTTAGTCTACAGGTGTACCCAAC
P69           ATCTCATATTAATCCTTGTGTGCCTGTTTGTCTGTTGTTAGTCTACAGGTGTACCCAAC
P24           ATCTCATATTAATCCTTGTGTGCCTGTTTGTCTGTTGTTAGTCTACAGGTGTACCCAAC
P23           ATCTCATATTAATCCTTGTGTGCCTGTTTGTCTGTTGTTAGTCTACAGGTGTACCCAAC
P76           ATCTCATATTAATCCTTGTGTGCCTGTTTGTCTGTTGTTAGTCTACAGGTGTACCCAAC
P28           ATCTCATATTAATCCTTGTGTGCCTGTTTGTCTGTTGTTAGTCTACAGGTGTACCCAAC
              *****

K113_env      AGCTCCGACGAGACAGCGACCATCGAGAATGGGCCATGATGACGATGGTGGTTTTGTCTGA
P74           AGCTCCGAAGAGACAGCGACCATCGAGAATGGGCCATGATGACGATGGTGGTTTTGTCTGA
P2            AGCTCCGAAGAGACAGCGACCATCGAGAATGGGCCATGATGACGATGGTGGTTTTGTCTGA
P69           AGCTCCGAAGAGACAGCGACCATCGAGAATGGGCCATGATGACGATGGTGGTTTTGTCTGA
P24           AGCTCCGAAGAGACAGCGACCATCGAGAATGGGCCATGATGACGATGGTGGTTTTGTCTGA
P23           AGCTCCGAAGAGACAGCGACCATCGAGAATGGGCCATGATGACGATGGTGGTTTTGTCTGA
P76           AGCTCCGAAGAGACAGCGACCATCGAGAATGGGCCATGATGACGATGGTGGTTTTGTCTGA
P28           AGCTCCGAAGAGACAGCGACCATCGAGAATGGGCCATGATGACGATGGTGGTTTTGTCTGA
              *****

K113_env      AAAGAAAAGGGGAAATGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGT
P74           AAAGAAAAGGGGAAATGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGT
P2            AAAGAAAAGGGGAAATGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGT
P69           AAAGAAAAGGGGAAATGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGT
P24           AAAGAAAAGGGGAAATGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGT
P23           AAAGAAAAGGGGAAATGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGT
P76           AAAGAAAAGGGGAAATGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGT
P28           AAAGAAAAGGGGAAATGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGT
              *****

K113_env      AGAAAGAAGTAGACATAGGAGACTCCATTTTGTATGTATTAAGAAAAATTC TTCTGCCT
P74           AGAAAGAAGTAGACATAGGAGACTCCATTTTGTATGTATTAAGAAAAATTC TTCTGCCT
P2            AGAAAGAAGTAGACATAGGAGACTCCATTTTGTATGTATTAAGAAAAATTC TTCTGCCT
P69           AGAAAGAAGTAGACATAGGAGACTCCATTTTGTATGTATTAAGAAAAATTC TTCTGCCT
P24           AGAAAGAAGTAGACATAGGAGACTCCATTTTGTATGTATTAAGAAAAATTC TTCTGCCT
P23           AGAAAGAAGTAGACATAGGAGACTCCATTTTGTATGTATTAAGAAAAATTC TTCTGCCT
P76           AGAAAGAAGTAGACATAGGAGACTCCGTTTGTATGTATTAAGAAAAATTC TTCTGCCT
P28           AGAAAGAAGTAGACATAGGAGACTCCATTTTGTATGTATTAAGAAAAATTC TTCTGCCT
              *****

K113_env      TGAGATTCTGTTAATCTATGACCTTACCCCAACCCCGTGCTCTCT-GAAACATCTGCTG
P74           TGAGATTCTGTTAATCTATGACCTTACCCCAACCCCGTGCTCTCT-GAAACATCTGCTG
P2            TGAGATTCTGTTAATCTATGACCTTACCCCAACCCCGTGCTCTCT-GAAACATCTGCTG
P69           TGAGATTCTGTTAATCTATGACCTTACCCCAACCCCGTGCTCTCT-GAAACATCTGCTG
P24           TGAGATTCTGTTAATCTATGACCTTACCCCAACCCCGTGCTCTCT-GAAACATCTGCTG
P23           TGAGATTCTGTTAATCTATGACCTTACCCCAACCCCGTGCTCTCT-GAAACATCTGCTG
P76           TGAGATTCTGTTAATCTATGACCTTACCCCAACCCCGTGCTCTCTGAAACATCTGCTG
P28           TGAGATTCTGTTAATCTATGACCTTACCCCAACCCCGTGCTCTCT-GAAACATCTGCTG
              *****

```

```

K113_env      TGTCAACTCAGAGTTGAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTAAACAGATGC
P74           TGTCAACTCAGAGTTGAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTAAACAGATGC
P2            TGTCAACTCAGAGTTGAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTAAACAGATGC
P69           TGTCAACTCAGAGTTGAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTAAACAGATGC
P24           TGTCAACTCAGAGTTGAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTAAACAGATGC
P23           TGTCAACTCAGAGTTGAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTAAACAGATGC
P76           TGTCAACTCAGAGTTGAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTAAACAGATGC
P28           TGTCAACTCAGAGTTGAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTAAACAGATGC
*****

K113_env      TTGAAGGCAGCATGCTCCTTAAGAGTCATCACCCTCCCTAATCTCAAGTACCCAGGGAC
P74           TTGAAGGCAGCATGCTCCTTAAGAGTCATCACCCTCCCTAATCTCAAGTACCCAGGGAC
P2            TTGAAGGCAGCATGCTCCTTAAGAGTCATCACCCTCCCTAATCTCAAGTACCCAGGGAC
P69           TTGAAGGCAGCATGCTCCTTAAGAGTCATCACCCTCCCTAATCTCAAGTACCCAGGGAC
P24           TTGAAGGCAGCATGCTCCTTAAGAGTCATCACCCTCCCTAATCTCAAGTACCCAGGGAC
P23           TTGAAGGCAGCATGCTCCTTAAGAGTCATCACCCTCCCTAATCTCAAGTACCCAGGGAC
P76           TTGAAGGCAGCATGCTCCTTAAGAGTCATCACCCTCCCTAATCTCAAGTACCCAGGGAC
P28           TTGAAGGCAGCATGCTCCTTAAGAGTCATCACCCTCCCTAATCTCAAGTACCCAGGGAC
*****

K113_env      ACAAAAACCTGCAGAAGGCCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGT
P74           ACAAAAACCTGCAGAAGGCCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGT
P2            ACAAAAACCTGCAGAAGGCCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGT
P69           ACAAAAACCTGCAGAAGGCCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGT
P24           ACAAAAACCTGCAGAAGGCCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGT
P23           ACAAAAACCTGCAGAAGGCCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGT
P76           ACAAAAACCTGCAGAAGGCCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGT
P28           ACAAAAACCTGCAGAAGGCCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGT
*****

K113_env      TTCTCCCCATGTGATAGTCTGAAATATGGCCTCGTGGAAGGAAAGACCTGACCGTCCC
P74           TTCTCCCCATGTGATAGTCTGAAATATGGCCTCGTGGAAGGAAAGACCTGACCGTCCC
P2            TTCTCCCCATGTGATAGTCTGAAATATGGCCTCGTGGAAGGAAAGACCTGACCGTCCC
P69           TTCTCCCCATGTGATAGTCTGAAATATGGCCTCGTGGAAGGAAAGACCTGACCGTCCC
P24           TTCTCCCCATGTGATAGTCTGAAATATGGCCTCGTGGAAGGAAAGACCTGACCGTCCC
P23           TTCTCCCCATGTGATAGTCTGAAATATGGCCTCGTGGAAGGAAAGACCTGACCGTCCC
P76           TTCTCCCCATGTGATAGTCTGAAATATGGCCTCGTGGAAGGAAAGACCTGACCGTCCC
P28           TTCTCCCCATGTGATAGTCTGAAATATGGCCTCGTGGAAGGAAAGACCTGACCGTCCC
*****

K113_env      CCAGCCCGACACCCGTAAAGGGTCTGTGCTGAGGAGGATTAGTATAAGAGGAAGGAATGC
P74           CCAGCCCGACACCTGTAAAGGGTCTGTGCTGAGGAGGATTAGTAAAAGAGGAAGGAATGC
P2            CCAGCCCGACACCTGTAAAGGGTCTGTGCTGAGGAGGATTAGTAAAAGAGGAAGGAATGC
P69           CCAGCCCGACACCTGTAAAGGGTCTGTGCTGAGGAGGATTAGTAAAAGAGGAAGGAATGC
P24           CCAGCCCGACACCTGTAAAGGGTCTGTGCTGAGGAGGATTAGTAAAAGAGGAAGGAATGC
P23           CCAGCCCGACACCTGTAAAGGGTCTGTGCTGAGGAGGATTAGTAAAAGAGGAAGGAATGC
P76           CCAGCCCGACACCTGTAAAGGGTCTGTGCTGAGGAGGATTAGTAAAAGAGGAAGGAATGC
P28           CCAGCCCGACACCTGTAAAGGGTCTGTGCTGAGGAGGATTAGTAAAAGAGGAAGGAATGC
*****

K113_env      CTCTTGCAAGTTGAGACAAGAGGAAGGCATCTGTCTCCTCCCTGTCCCTGGGCAATGGAAT
P74           CTCTTGCAAGTTGAGACAAGAGGAAGGCATCTGTCTCCTCCCTGTCCCTGGGCAATGGAAT
P2            CTCTTGCAAGTTGAGACAAGAGGAAGGCATCTGTCTCCTCCCTGTCCCTGGGCAATGGAAT
P69           CTCTTGCAAGTTGAGACAAGAGGAAGGCATCTGTCTCCTCCCTGTCCCTGGGCAATGGAAT
P24           CTCTTGCAAGTTGAGACAAGAGGAAGGCATCTGTCTCCTCCCTGTCCCTGGGCAATGGAAT
P23           CTCTTGCAAGTTGAGACAAGAGGAAGGCATCTGTCTCCTCCCTGTCCCTGGGCAATGGAAT
P76           CTCTTGCAAGTTGAGACAAGAGGAAGGCATCTGTCTCCTCCCTGTCCCTGGGCAATGGAAT
P28           CTCTTGCAAGTTGAGACAAGAGGAAGGCATCTGTCTCCTCCCTGTCCCTGGGCAATGGAAT
*****

K113_env      GTCTCGGTATAAAACCCGATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAG
P74           GTCTCGGTATAAAACCCGATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAG
P2            GTCTCGGTATAAAACCCGATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAG
P69           GTCTCGGTATAAAACCCGATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAG
P24           GTCTCGGTATAAAACCCGATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAG
P23           GTCTCGGTATAAAACCCGATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAG
P76           GTCTCGGTATAAAACCCGATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAG
P28           GTCTCGGTATAAAACCCGATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAG
*****

```

```

K113_env      GGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCATTGAGATGTTTATGT
P74           GGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCATTGAGATGTTTATGT
P2            GGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCATTGAGATGTTTATGT
P69           GGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCATTGAGATGTTTATGT
P24           GGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCATTGAGATGTTTATGT
P23           GGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCATTGAGATGTTTATGT
P76           GGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCATTGAGATGTTTATGT
P28           GGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCATTGAGATGTTTATGT
*****

K113_env      GTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCCAAGACCTT
P74           GTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCCAAGACCTT
P2            GTATGCATATCTAAAAGCACAGCACTTAATCCTTTATATTGTCTATGATGCCAAGACCTT
P69           GTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCCAAGACCTT
P24           GTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCCAAGACCTT
P23           GTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCCAAGACCTT
P76           GTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCCAAGACCTT
P28           GTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCCAAGACCTT
*****

K113_env      TGTTACAGTGTTTGTCTGCTGACCCCTCTCCCCACAATTGTCTTGTGACCCCTGACACATCC
P74           TGTTACAGTGTTTGTCTGCTGACCCCTCTCCCCACAATTGTCTTGTGACCCCTGACACATCC
P2            TGTTACAGTGTTTGTCTGCTGACCCCTCTCCCCACAATTGTCTTGTGACCCCTGACACATCC
P69           TGTTACAGTGTTTGTCTGCTGACCCCTCTCCCCACAATTGTCTTGTGACCCCTGACACATCC
P24           TGTTACAGTGTTTGTCTGCTGACCCCTCTCCCCACAATTGTCTTGTGACCCCTGACACATCC
P23           TGTTACAGTGTTTGTCTGCTGACCCCTCTCCCCACAATTGTCTTGTGACCCCTGACACATCC
P76           TGTTACAGTGTTTGTCTGCTGACCCCTCTCCCCACAATTGTCTTGTGACCCCTGACACATCC
P28           TGTTACAGTGTTTGTCTGCTGACCCCTCTCCCCACAATTGTCTTGTGACCCCTGACACATCC
*****

K113_env      CCCTCTTTGAGAAACACCCACAGATGATCAATAAATACTAAGGGAACCTAGAGGCTGGCG
P74           CCCTCTTTGAGAAACACCCACAGATGATCAATAAATACTAAGGGAACCTAGAGGCTGGCG
P2            CCCTCTTTGAGAAACACCCACAGATGATCAATAAATACTAAGGGAACCTAGAGGCTGGCG
P69           CCCTCTTTGAGAAACACCCACAGATGATCAATAAATACTAAGGGAACCTAGAGGCTGGCG
P24           CCCTCTTTGAGAAACACCCACAGATGATCAATAAATACTAAGGGAACCTAGAGGCTGGCG
P23           CCCTCTTTGAGAAACACCCACAGATGATCAATAAATACTAAGGGAACCTAGAGGCTGGCG
P76           CCCTCTTTGAGAAACACCCACAGATGATCAATAAATACTAAGGGAACCTAGAGGCTGGCG
P28           CCCTCTTTGAGAAACACCCACAGATGATCAATAAATACTAAGGGAACCTAGAGGCTGGCG
*****

K113_env      GGATCCCTCCATATGCTGAACGCTGGTTCCCCGGTTCCCCTTATTTCTTTCTCTATACCTT
P74           GGATCCCTCCATATGCTGAACGCTGGTTCCCCGGTTCCCCTTACTTCTTTCTCTATACCTT
P2            GGATCCCTCCATATGCTGAACGCTGGTTCCCCGGTTCCCCTTATTTCTTTCTCTATACCTT
P69           GGATCCCTCCATATGCTGAACGCTGGTTCCCCGGTTCCCCTTATTTCTTTCTCTATACCTT
P24           GGATCCCTCCATATGCTGAACGCTGGTTCCCCGGTTCCCCTTATTTCTTTCTCTATACCTT
P23           GGATCCCTCCATATGCTGAACGCTGGTTCCCCGGTTCCCCTTATTTCTTTCTCTATACCTT
P76           GGATCCCTCCATATGCTGAACGCTGGTTCCCCGGTTCCCCTTATTTCTTTCTCTATACCTT
P28           GGATCCCTCCATATGCTGAACGCTGGTTCCCCGGTTCCCCTTATTTCTTTCTCTATACCTT
*****

K113_env      GTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGG
P74           GTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGG
P2            GTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGG
P69           GTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGG
P24           GTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGG
P23           GTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGG
P76           GTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGG
P28           GTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGG
*****

K113_env      TGTGTAGGGGCAACCCACCCCTACACCTCTATAATTTTCTTACACCTAAGGTTTATCTTTA
P74           TGTGTAGGGGCAACCCACCCCTACACCTCTATAATTTTCTTACACCTAAGGTTTATCTTTA
P2            TGTGTAGGGGCAACCCACCCCTACACCTCTATAATTTTCTTACACCTAAGGTTTATCTTTA
P69           TGTGTAGGGGCAACCCACCCCTACACCTCTATAATTTTCTTACACCTAAGGTTTATCTTTA
P24           TGTGTAGGGGCAACCCACCCCTACACCTCTATAATTTTCTTACACCTAAGGTTTATCTTTA
P23           TGTGTAGGGGCAACCCACCCCTACACCTCTATAATTTTCTTACACCTAAGGTTTATCTTTA
P76           TGTGTAGGGGCAACCCACCCCTACACCTCTATAATTTTCTTGCACCTAAGGTTTATCTTTA
P28           TGTGTAGGGGCAACCCACCCCTACACCTCTATAATTTTCTTACACCTAAGGTTTATCTTTA
*****

```

```

K113_env -----
P74 GACTAACATATATTGAACTCTATGTAAATCAAACTAAACGCTGTCATGTGTA-
P2 GACTAACATATATTGAACTCTATGTAAATCAAACTAAACGCTGTCATGTGTA-
P69 GACTAACATATATTGAACTCTATGTAAATCAAACTAAACGCTGTCATGTAATC
P24 GACTAACATATATTGAACTCTATGTAAATCAAACTAAACGCTGTCATGTGTA-
P23 GACTAACATATATTGAACTCTATGTAAATCAAACTAAACGCTGTCATGTGT--
P76 GACTAACATATATTGAACTCTATGTAAATCAAACTAAACGCTGTCATGTGTA-
P28 GACTAACATATATTGAACTCTATGTAAATCAAACTANNNNNGTAA-----

```

### 7.3 Das sequenzierte *gag*-Gen

```

P24 AAAAGCCATCGATTGTACAGTAATAAGTTCAATACAAAGCAGAGAATATACTTGCTTTCA
P76 AAAAGCCATCGATTGTACAGTAATAAGTTCAATACAAAGCAGAGAATATACTTGCTTTCA
P73 AAAAGCCATCGATTGTACAGTAATAAGTTCAATACAAAGCAGAGAATATACTTGCTTTCA
P74 AAAAGCCATCGATTGTACAGTAATAAGTTCAATACAAAGCAGAGAATATACTTGCTTTCA
P69 -AAAGCCATCGATTGTACAGTAATAAGTTCAATACAAAGCAGAGAATATACTTGCTTTCA
K113_gag -----

```

```

P24 GCATTTTAAAGGTTTTAGTTTTCTAGTAGTCATCTTGTGAAATGATTTGCCTTGTTCA
P76 GCATTTTAAAGGTTTTAGTTTTCTAGTAGTCATCTTGTGAAATGATTTGCCTTGTTCA
P73 GCATTTTAAAGGTTTTAGTTTTCTAGTAGTCATCTTGTGAAATGATTTGCCTTGTTCA
P74 GCATTTTAAAGGTTTTAGTTTTCTAGTAGTCATCTTGTGAAATGATTTGCCTTGTTCA
P69 GCATTTTAAAGGTTTTAGTTTTCTAGTAGTCATCTTGTGAAATGATTTGCCTTGTTCA
K113_gag -----

```

```

P24 ATATTGTTTTCTTCTGAAAATAGTTACAAATTCATACTTACACAAACTCACTTACTCTAT
P76 ATATTGTTTTCTTCTGAAAATAGTTACAAATTCATACTTACACAAACTCACTTACTCTAT
P73 ATATTGTTTTCTTCTGAAAATAGTTACAAATTCATACTTACACAAACTCACTTACTCTAT
P74 ATATTGTTTTCTTCTGAAAATAGTTACAAATTCATACTTACACAAACTCACTTACTCTAT
P69 ATATTGTTTTCTTCTGAAAATAGTTACAAATTCATACTTACACAAACTCACTTACTCTAT
K113_gag -----

```

```

P24 TGTGGGAAAAAGCAAGAGAGATCAGATTGTTACTGTGCTCTGTGTAGAAAGAAGTAGACAT
P76 TGTGGGAAAAAGCAAGAGAGATCAGATTGTTACTGTGCTCTGTGTAGAAAGAAGTAGACAT
P73 TGTGGGAAAAAGCAAGAGAGATCAGATTGTTACTGTGCTCTGTGTAGAAAGAAGTAGACAT
P74 TGTGGGAAAAAGCAAGAGAGATCAGATTGTTACTGTGCTCTGTGTAGAAAGAAGTAGACAT
P69 TGTGGGAAAAAGCAAGAGAGATCAGATTGTTACTGTGCTCTGTGTAGAAAGAAGTAGACAT
K113_gag TGTGGGAAAAAGCAAGAGAGATCAGATTGTTACTGTGCTCTGTGTAGAAAGAAGTAGACAT
*****.*****

```

```

P24 AGGAGACTCCATTTTGTATGTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATC
P76 AGGAGACTCCATTTTGTATGTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATC
P73 AGGAGACTCCATTTTGTATGTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATC
P74 AGGAGACTCCATTTTGTATGTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATC
P69 AGGAGACTCCATTTTGTATGTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATC
K113_gag AGGAGACTCCATTTTGTATGTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATC
*****.*****

```

```

P24 TATGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTGCTGTGTCAACTCAAAGTTGA
P76 TATGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTGCTGTGTCAACTCAAAGTTGA
P73 TATGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTGCTGTGTCAACTCAAAGTTGA
P74 TATGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTGCTGTGTCAACTCAAAGTTGA
P69 TATGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTGCTGTGTCAACTCAAAGTTGA
K113_gag TATGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTGCTGTGTCAACTCAGAGTTGA
*****.*****

```

P24 ATGGATTAAAGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCCTC  
P76 ATGGATTAAAGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCCTC  
P73 ATGGATTAAAGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCCTC  
P74 ATGGATTAAAGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCCTC  
P69 ATGGATTAAAGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCCTC  
K113\_gag ATGGATTAAAGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCCTC  
\*\*\*\*\*

P24 CTTAAGAGTCATCACCCTCCCTAATCTCAAGTACCCAGGGACACAAAACTGCGGAAGG  
P76 CTTAAGAGTCATCACCCTCCCTAATCTCAAGTACCCAGGGACACAAAACTGCGGAAGG  
P73 CTTAAGAGTCATCACCCTCCCTAATCTCAAGTACCCAGGGACACAAAACTGCGGAAGG  
P74 CTTAAGAGTCATCACCCTCCCTAATCTCAAGTACCCAGGGACACAAAACTGCGGAAGG  
P69 CTTAAGAGTCATCACCCTCCCTAATCTCAAGTACCCAGGGACACAAAACTGCGGAAGG  
K113\_gag CTTAAGAGTCATCACCCTCCCTAATCTCAAGTACCCAGGGACACAAAACTGCGGAAGG  
\*\*\*\*\*

P24 CCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTCTCCCCATGTGATAG  
P76 CCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTCTCCCCATGTGATAG  
P73 CCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTCTCCCCATGTGATAG  
P74 CCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTCTCCCCATGTGATAG  
P69 CCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTCTCCCCATGTGATAG  
K113\_gag CCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTCTCCCCATGTGATAG  
\*\*\*\*\*

P24 TCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGTCCCCAGCCCGACACCCGTA  
P76 TCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGTCCCCAGCCCGACACCCGTA  
P73 TCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGTCCCCAGCCCGACACCCGTA  
P74 TCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGTCCCCAGCCCGACACCCGTA  
P69 TCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGTCCCCAGCCCGACACCCGTA  
K113\_gag TCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGTCCCCAGCCCGACACCCGTA  
\*\*\*\*\*

P24 AAGGGTCTGTGCTGAGGAGGATTAGTATAAGAGGAAGGAATGCCTCTTGCAGTTGAGACA  
P76 AAGGGTCTGTGCTGAGGAGGATTAGTATAAGAGGAAGGAATGCCTCTTGCAGTTGAGACA  
P73 AAGGGTCTGTGCTGAGGAGGATTAGTATAAGAGGAAGGAATGCCTCTTGCAGTTGAGACA  
P74 AAGGGTCTGTGCTGAGGAGGATTAGTATAAGAGGAAGGAATGCCTCTTGCAGTTGAGACA  
P69 AAGGGTCTGTGCTGAGGAGGATTAGTATAAGAGGAAGGAATGCCTCTTGCAGTTGAGACA  
K113\_gag AAGGGTCTGTGCTGAGGAGGATTAGTATAAGAGGAAGGAATGCCTCTTGCAGTTGAGACA  
\*\*\*\*\*

P24 AGAGGAAGGCATCTGTCTCCTCCCTGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCC  
P76 AGAGGAAGGCATCTGTCTCCTCCCTGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCC  
P73 AGAGGAAGGCATCTGTCTCCTCCCTGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCC  
P74 AGAGGAAGGCATCTGTCTCCTCCCTGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCC  
P69 AGAGGAAGGCATCTGTCTCCTCCCTGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCC  
K113\_gag AGAGGAAGGCATCTGTCTCCTCCCTGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCC  
\*\*\*\*\*

P24 GATTGTATGCTCCATCTACTGAGACAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCT  
P76 GATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCT  
P73 GATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCT  
P74 GATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCT  
P69 GATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCT  
K113\_gag GATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCT  
\*\*\*\*\*

P24 GCGGGCAGCAATACTGCTTTGTAAAGCACTGAGATGTTTATGTGTATGCATATCTAAAAG  
P76 GCGGGCAGCAATACTGCTTTGTAAAGCACTGAGATGTTTATGTGTATGCATATCTAAAAG  
P73 GCGGGCAGCAATACTGCTTTGTAAAGCACTGAGATGTTTATGTGTATGCATATCTAAAAG  
P74 GCGGGCAGCAATACTGCTTTGTAAAGCACTGAGATGTTTATGTGTATGCATATCTAAAAG  
P69 GCGGGCAGCAATACTGCTTTGTAAAGCACTGAGATGTTTATGTGTATGCATATCTAAAAG  
K113\_gag GCGGGCAGCAATACTGCTTTGTAAAGCACTGAGATGTTTATGTGTATGCATATCTAAAAG  
\*\*\*\*\*

P24 CACAGCACTTAATCCTTTACATTGTCTATGATGCCAAGACCTTTGTTACCGTGTGTTGTCT  
P76 CACAGCACTTAATCCTTTACATTGTCTATGATGCCAAGACCTTTGTTACCGTGTGTTGTCT  
P73 CACAGCACTTAATCCTTTACATTGTCTATGATGCCAAGACCTTTGTTACCGTGTGTTGTCT  
P74 CACAGCACTTAATCCTTTACATTGTCTATGATGCCAAGACCTTTGTTACCGTGTGTTGTCT  
P69 CACAGCACTTAATCCTTTACATTGTCTATGATGCCAAGACCTTTGTTACCGTGTGTTGTCT  
K113\_gag CACAGCACTTAATCCTTTACATTGTCTATGATGCCAAGACCTTTGTTACCGTGTGTTGTCT  
\*\*\*\*\*



P24 GCTGACCCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTTGAGAAACAC  
P76 GCTGACCCCTCTCCCCACAATTGTCTTATGACCCTGACACATCCCCCTCTTTGAGAAACAC  
P73 GCTGACCCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTTGAGAAACAC  
P74 GCTGACCCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTTGAGAAACAC  
P69 GCTGACCCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTTGAGAAACAC  
K113\_gag GCTGACCCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTTGAGAAACAC  
\*\*\*\*\*

P24 CCACAGATGATCAATAAATACTAAGGGAACCTGAGAGGCTGGCGGGATCCTCCATATGCTG  
P76 CCACAGATGATCAATAAATACTAAGGGAACCTGAGAGGCTGGCGGGATCCTCCATATGCTG  
P73 CCACAGATGATCAATAAATACTAAGGGAACCTGAGAGGCTGGCGGGATCCTCCATATGCTG  
P74 CCACAGATGATCAATAAATACTAAGGGAACCTGAGAGGCTGGCGGGATCCTCCATATGCTG  
P69 CCACAGATGATCAATAAATACTAAGGGAACCTGAGAGGCTGGCGGGATCCTCCATATGCTG  
K113\_gag CCACAGATGATCAATAAATACTAAGGGAACCTGAGAGGCTGGCGGGATCCTCCATATGCTG  
\*\*\*\*\*

P24 AACGCTGGTTCCCCGGTTCCCTTATTTCTTTCTCTATACTTTGTCTCTGTGTCTTTTTTC  
P76 AACGCTGGTTCCCCGGTTCCCTTATTTCTTTCTCTATACTTTGTCTCTGTGTCTTTTTTC  
P73 AACGCTGGTTCCCCGGTTCCCTTATTTCTTTCTCTATACTTTGTCTCTGTGTCTTTTTTC  
P74 AACGCTGGTTCCCCGGTTCCCTTATTTCTTTCTCTATACTTTGTCTCTGTGTCTTTTTTC  
P69 AACGCTGGTTCCCCGGTTCCCTTATTTCTTTCTCTATACTTTGTCTCTGTGTCTTTTTTC  
K113\_gag AACGCTGGTTCCCCGGTTCCCTTATTTCTTTCTCTATACTTTGTCTCTGTGTCTTTTTTC  
\*\*\*\*\*

P24 TTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAACCCA  
P76 TTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAACCCA  
P73 TTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAACCCA  
P74 TTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAACCCA  
P69 TTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAACCCA  
K113\_gag TTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAACCCA  
\*\*\*\*\*

P24 CCCCTACATCTGGTGCCCAACGTGGAGGTTTTTCTCTAGGGTGAAGGTACGCTCGAGCGT  
P76 CCCCTACATCTGGTGCCCAACGTGGAGGTTTTTCTCTAGGGTGAAGGTACGCTCGAGCGT  
P73 CCCCTACATCTGGTGCCCAACGTGGAGGTTTTTCTCTAGGGTGAAGGTACGCTCGAGCGT  
P74 CCCCTACATCTGGTGCCCAACGTGGAGGTTTTTCTCTAGGGTGAAGGTACGCTCGAGCGT  
P69 CCCCTACATCTGGTGCCCAACGTGGAGGTTTTTCTCTAGGGTGAAGGTACGCTCGAGCGT  
K113\_gag CCCCTACATCTGGTGCCCAACGTGGAGGTTTTTCTCTAGGGTGAAGGTACGCTCGAGCGT  
\*\*\*\*\*

P24 GGTCATTGAGGACAAGTCGACGAGAGATCCCGAGTACGTCTACAGTCAGCCTTATGGTAA  
P76 GGTCATTGAGGACAAGTCGACGAGAGATCCCGAGTACGTCTACAGTCAGCCTTATGGTAA  
P73 GGTCATTGAGGACAAGTCGACGAGAGATCCCGAGTACGTCTACAGTCAGCCTTATGGTAA  
P74 GGTCATTGAGGACAAGTCGACGAGAGATCCCGAGTACGTCTACAGTCAGCCTTATGGTAA  
P69 GGTCATTGAGGACAAGTCGACGAGAGATCCCGAGTACGTCTACAGTCAGCCTTATGGTAA  
K113\_gag GGTCATTGAGGACAAGTCGACGAGAGATCCCGAGTACGTCTACAGTCAGCCTTATGGTAA  
\*\*\*\*\*

P24 GCTTGTGCGCTCAGAAGAAGCTAGGGTGATAATGGGGCAAACATAAAGTAAAATTAAAAG  
P76 GCTTGTGCGCTCAGAAGAAGCTAGGGTGATAATGGGGCAAACATAAAGTAAAATTAAAAG  
P73 GCTTGTGCGCTCAGAAGAAGCTAGGGTGATAATGGGGCAAACATAAAGTAAAATTAAAAG  
P74 GCTTGTGCGCTCAGAAGAAGCTAGGGTGATAATGGGGCAAACATAAAGTAAAATTAAAAG  
P69 GCTTGTGCGCTCAGAAGAAGCTAGGGTGATAATGGGGCAAACATAAAGTAAAATTAAAAG  
K113\_gag GCTTGTGCGCTCAGAAGAAGCTAGGGTGATAATGGGGCAAACATAAAGTAAAATTAAAAG  
\*\*\*\*\*

P24 TAAATATGCCTCTTATCTCAGCTTTATTAATAATCTTTTAAAAAGAGGGGGAGTTAAAGT  
P76 TAAATATGCCTCTTATCTCAGCTTTATTAATAATCTTTTAAAAAGAGGGGGAGTTAAAGT  
P73 TAAATATGCCTCTTATCTCAGCTTTATTAATAATCTTTTAAAAAGAGGGGGAGTTAAAGT  
P74 TAAATATGCCTCTTATCTCAGCTTTATTAATAATCTTTTAAAAAGAGGGGGAGTTAAAGT  
P69 TAAATATGCCTCTTATCTCAGCTTTATTAATAATCTTTTAAAAAGAGGGGGAGTTAAAGT  
K113\_gag TAAATATGCCTCTTATCTCAGCTTTATTAATAATCTTTTAAAAAGAGGGGGAGTTAAAGT  
\*\*\*\*\*

P24 ATCTACAAAAATCTAATCAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCC  
P76 ATCTACAAAAATCTAATCAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCC  
P73 ATCTACAAAAATCTAATCAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCC  
P74 ATCTACAAAAATCTAATCAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCC  
P69 ATCTACAAAAATCTAATCAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCC  
K113\_gag ATCTACAAAAATCTAATCAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCC  
\*\*\*\*\*

P24 AGAACAAGGAACCTTTAGATCTAAAAGACTGGAAAAGAATTGGTAAGGAACTAAAAACAAGC  
P76 AGAACAAGGAACCTTTAGATCTAAAAGACTGGAAAAGAATTGGTAAGGAACTAAAAACAAGC  
P73 ARAACAAGGAACCTTTAGATCTAAAAGACTGGAAAAGAATTGGTAAGGAACTAAAAACAAGC  
P74 AGAACAAGGAACCTTTAGATCTAAAAGACTGGAAAAGAATTGGTAAGGAACTAAAAACAAGC  
P69 AGAACAAGGAACCTTTAGATCTAAAAGACTGGAAAAGAATTGGTAAGGAACTAAAAACAAGC  
K113\_gag AGAACAAGGAACCTTTAGATCTAAAAGACTGGAAAAGAATTGGTAAGGAACTAAAAACAAGC  
\* \* \* \* \*

P24 AGGTAGGAAGGGTAATATCATTTCCACTTACAGTATGGAATGATTGGGCCATTATTAAAGC  
P76 AGGTAGGAAGGGTAATATCATTTCCACTTACAGTATGGAATGATTGGGCCATTATTAAAGC  
P73 AGGTAGGAAGGGTAATATCATTTCCACTTACAGTACGGAATGATTGGGCCATTATTAAAGC  
P74 AGGTAGGAAGGGTAATATCATTTCCACTTACAGTATGGAATGATTGGGCCATTATTAAAGC  
P69 AGGTAGGAAGGGTAATATCATTTCCACTTACAGTATGGAATGATTGGGCCATTATTAAAGC  
K113\_gag AGGTAGGAAGGGTAATATCATTTCCACTTACAGTATGGAATGATTGGGCCATTATTAAAGC  
\* \* \* \* \*

P24 AGCTTTAGAACCATTTCAAACAGAAGAAGATAGTGTTTCAGTTTCTGATGCCCTTGGAAAG  
P76 AGCTTTAGAACCATTTCAAACAGAAGAAGATAGTGTTTCAGTTTCTGATGCCCTTGGAAAG  
P73 AGCTTTAGAACCATTTCAAACAGAAAAAATAGTGTTTCAGTTCCTGATGCCCTTGGAAAG  
P74 AGCTTTAGAACCATTTCAAACAGAAGAAGATAGTGTTTCAGTTTCTGATGCCCTTGGAAAG  
P69 AGCTTTAGAACCATTTCAAACAGAAGAAGATAGTGTTTCAGTTTCTGATGCCCTTGGAAAG  
K113\_gag AGCTTTAGAACCATTTCAAACAGAAGAAGATAGTGTTTCAGTTTCTGATGCCCTTGGAAAG  
\* \* \* \* \*

P24 CTGTATAATAGATTGTAATGAAAAGACAAGGAAAAATCCCAGAAAGAAACCGAAAGTTT  
P76 CTGTATAATAGATTGTAATGAAAAGACAAGGAAAAATCCCAGAAAGAAACCGAAAGTTT  
P73 CTGTATAATAAATTGTAATGAAAAGACAAGGAAAAATCCCAGAAAGAAACCGAAAGTTT  
P74 CTGTATAATAGATTGTAATGAAAAGACAAGGAAAAATCCCAGAAAGAAACCGAAAGTTT  
P69 CTGTATAATAGATTGTAATGAAAAGACAAGGAAAAATCCCAGAAAGAAACCGAAAGTTT  
K113\_gag CTGTATAATAGATTGTAATGAAAAGACAAGGAAAAATCCCAGAAAGAAACCGAAAGTTT  
\* \* \* \* \*

P24 ACATTGCGAATATGTAGCAGAGCCAGTAATGGCTCAGTCAACGCAAAATGTTGACTATAA  
P76 ACATTGCGAATATGTAGCAGAGCCAGTAATGGCTCAGTCAACGCAAAATGTTGACTATAA  
P73 ACATTGCGAATATGTAGCAAGCCAGTAATGGCTCAGTCAACGCAAAATGTTGACTATAA  
P74 ACATTGCGAATATGTAGCAGAGCCAGTAATGGCTCAGTCAACGCAAAATGTTGACTATAA  
P69 ACATTGCGAATATGTAGCAGAGCCAGTAATGGCTCAGTCAACGCAAAATGTTGACTATAA  
K113\_gag ACATTGCGAATATGTAGCAGAGCCAGTAATGGCTCAGTCAACGCAAAATGTTGACTATAA  
\* \* \* \* \*

P24 TCAATTACAGGAGGTGATATATCCTGAAACGTTAAAAATTAGAAGGAAAAGGTCCAGAATT  
P76 TCAATTACAGGAGGTGATATATCCTGAAACGTTAAAAATTAGAAGGAAAAGGTCCAGAATT  
P73 TCAATTACAGGAGGTGATATATCCTGAAACGTTAAAAATTAGAAGGAAAAGGTCCAGAATT  
P74 TCAATTACAGGAGGTGATATATCCTGAAACGTTAAAAATTAGAAGGAAAAGGTCCAGAATT  
P69 TCAATTACAGGAGGTGATATATCCTGAAACGTTAAAAATTAGAAGGAAAAGGTCCAGAATT  
K113\_gag TCAATTACAGGAGGTGATATATCCTGAAACGTTAAAAATTAGAAGGAAAAGGTCCAGAATT  
\* \* \* \* \*

P24 AATGGGGCCATCAGAGTCTAAACCACGAGGGCCAAGTCTCTTCCAGCAGGTGAGGTGCC  
P76 AATGGGGCCATCAGAGTCTAAACCACGAGGGCCAAGTCTCTTCCAGCAGGTGAGGTGCC  
P73 AATGGGGCCATCAAAAGTCTAAACCACGAGGGCCAAGTCTCTTCCAGCAGGTGAGGTGCC  
P74 AATGGGGCCATCAGAGTCTAAACCACGAGGGCCAAGTCTCTTCCAGCAGGTGAGGTGCC  
P69 AATGGGGCCATCAGAGTCTAAACCACGAGGGCCAAGTCTCTTCCAGCAGGTGAGGTGCC  
K113\_gag AATGGGGCCATCAGAGTCTAAACCACGAGGGCCAAGTCTCTTCCAGCAGGTGAGGTGCC  
\* \* \* \* \*

P24 CGTAACATTACAACCTCAAAGCAGGTTAAAGAAAATAAGACCCAACCGCCAGTAGCCTA  
P76 CGTAACATTACAACCTCAAAGCAGGTTAAAGAAAATAAGACCCAACCGCCAGTAGCCTA  
P73 CGTAACATTACAACCTCAAAGCAGGTTAAAGAAAATAAGACCCAACCGCCAGTAGCCTA  
P74 CGTAACATTACAACCTCAAAGCAGGTTAAAGAAAATAAGACCCAACCGCCAGTAGCCTA  
P69 CGTAACATTACAACCTCAAAGCAGGTTAAAGAAAATAAGACCCAACCGCCAGTAGCCTA  
K113\_gag CGTAACATTACAACCTCAAAGCAGGTTAAAGAAAATAAGACCCAACCGCCAGTAGCCTA  
\* \* \* \* \*

P24 CCAATACTGGCCGCGGCTGAACTTCAGTATCAGCCACCCCCAGAAAGTCAGTATGGATA  
P76 CCAATACTGGCCGCGGCTGAACTTCAGTATCAGCCACCCCCAGAAAGTCAGTATGGATA  
P73 CCAATACTGGCCGCGGCTGAACTTCAGTATCAGCCACCCCCAGAAAGTCAGTATGGATA  
P74 CCAATACTGGCCGCGGCTGAACTTCAGTATCAGCCACCCCCAGAAAGTCAGTATGGATA  
P69 CCAATACTGGCCGCGGCTGAACTTCAGTATCAGCCACCCCCAGAAAGTCAGTATGGATA  
K113\_gag CCAATACTGGCCGCGGCTGAACTTCAGTATCAGCCACCCCCAGAAAGTCAGTATGGATA  
\* \* \* \* \*

P24 TCCAGGAATGCCCCAGCACCACAGGGCAGGGCGCCATACCCTCAGCCGCCCTAGGAG  
P76 TCCAGGAATGCCCCAGCACCACAGGGCAGGGCGCCATACCCTCAGCCGCCCTAGGAG  
P73 TCCAGGAATGCCCCAGCACCACAGGGCAGGGCGCCATACCCTCAGCCGCCCTAGGAG  
P74 TCCAGGAATGCCCCAGCACCACAGGGCAGGGCGCCATACCCTCAGCCGCCCTAGGAG  
P69 TCCAGGAATGCCCCAGCACCACAGGGCAGGGCGCCATACCCTCAGCCGCCCTAGGAG  
K113\_gag TCCAGGAATGCCCCAGCACCACAGGGCAGGGCGCCATACCCTCAGCCGCCCTAGGAG  
\*\*\*\*\* . \*\*

P24 ACTTAATCCTACGGCACCACCTAGTAGACAGGGTAGTGAATTACATGAAATTATTGATAA  
P76 ACTTAATCCTACGGCACCACCTAGTAGACAGGGTAGTGAATTACATGAAATTATTGATAA  
P73 ACTTAATCCTACGGCACCACCTAGTAGACAGGGTAGTGAATTACATGAAATTATTGATAA  
P74 ACTTAATCCTACGGCACCACCTAGTAGACAGGGTAGTGAATTACATGAAATTATTGATAA  
P69 ACTTAATCCTACGGCACCACCTAGTAGACAGGGTAGTGAATTACATGAAATTATTGATAA  
K113\_gag ACTTAATCCTACGGCACCACCTAGTAGACAGGGTAGTGAATTACATGAAATTATTGATAA  
\*\*\*\*\*

P24 ATCAAGAAAGGAAGGAGATACTGAGGCATGGCAATTCCCAGTAACGTTAGAAGTATGATGCC  
P76 ATCAAGAAAGGAAGGAGATACTGAGGCATGGCAATTCCCAGTAACGTTAGAAGTATGATGCC  
P73 ATCAAGAAAGGAAGGAGATACTGAGGCATGGCAATTCCCAGTAACGTTAGAAGTATGATGCC  
P74 ATCAAGAAAGGAAGGAGATACTGAGGCATGGCAATTCCCAGTAACGTTAGAAGTATGATGCC  
P69 ATCAAGAAAGGAAGGAGATACTGAGGCATGGCAATTCCCAGTAACGTTAGAAGTATGATGCC  
K113\_gag ATCAAGAAAGGAAGGAGATACTGAGGCATGGCAATTCCCAGTAACGTTAGAAGTATGATGCC  
\*\*\*\*\*

P24 ACCTGGAGAAGGAGCCCAAGAGGGAGAGCCTCCACAGTTGAGGCCAGATACAAGTCTTT  
P76 ACCTGGAGAAGGAGCCCAAGAGGGAGAGCCTCCACAGTTGAGGCCAGATACAAGTCTTT  
P73 ACCTGGAGAAGGAGCCCAAGAGGGAGAGCCTCCACAGTTGAGGCCAGATACAAGTCTTT  
P74 ACCTGGAGAAGGAGCCCAAGAGGGAGAGCCTCCACAGTTGAGGCCAGATACAAGTCTTT  
P69 ACCTGGAGAAGGAGCCCAAGAGGGAGAGCCTCCACAGTTGAGGCCAGATACAAGTCTTT  
K113\_gag ACCTGGAGAAGGAGCCCAAGAGGGAGAGCCTCCACAGTTGAGGCCAGATACAAGTCTTT  
\*\*\*\*\*

P24 TTCAATAAAAAATGCTAAAAGATATGAAAGAGGGAGTAAAACAGTATGGACCCAACTCCCC  
P76 TTCAATAAAAAATGCTAAAAGATATGAAAGAGGGAGTAAAACAGTATGGACCCAACTCCCC  
P73 TTCAATAAAAAATGCTAAAAGATATGAAAGAGGGAGTAAAACAGTATGGACCCAACTCCCC  
P74 TTCAATAAAAAATGCTAAAAGATATGAAAGAGGGAGTAAAACAGTATGGACCCAACTCCCC  
P69 TTCAATAAAAAATGCTAAAAGATATGAAAGAGGGAGTAAAACAGTATGGACCCAACTCCCC  
K113\_gag TTCAATAAAAAATGCTAAAAGATATGAAAGAGGGAGTAAAACAGTATGGACCCAACTCCCC  
\*\*\* . \*\*\*\*\*

P24 TTATATGAGGACATTATTAGATTCCATTGCTCATGGACATAGACTCATTCCTTATGATTG  
P76 TTATATGAGGACATTATTAGATTCCATTGCTCATGGACATAGACTCATTCCTTATGATTG  
P73 TTATATGAGGACATTATTAGATTCCATTGCTCATGGACATAGACTCATTCCTTATGATTG  
P74 TTATATGAGGACATTATTAGATTCCATTGCTCATGGACATAGACTCATTCCTTATGATTG  
P69 TTATATGAGGACATTATTAGATTCCATTGCTCATGGACATAGACTCATTCCTTATGATTG  
K113\_gag TTATATGAGGACATTATTAGATTCCATTGCTCATGGACATAGACTCATTCCTTATGATTG  
\*\*\*\*\* . \*\*\*\*\*

P24 GGAGATTCTGGCAAAATCGTCTCTCTCACCCTCTCAATTTTACAATTTAAGACTTGGTG  
P76 GGAGATTCTGGCAAAATCGTCTCTCTCACCCTCTCAATTTTACAATTTAAGACTTGGTG  
P73 GGAGATTCTGGCAAAATCGTCTCTCTCACCCTCTCAATTTTACAATTTAAGACTTGGTG  
P74 GGAGATTCTGGCAAAATCGTCTCTCTCACCCTCTCAATTTTACAATTTAAGACTTGGTG  
P69 GGAGATTCTGGCAAAATCGTCTCTCTCACCCTCTCAATTTTACAATTTAAGACTTGGTG  
K113\_gag GGAGATTCTGGCAAAATCGTCTCTCTCACCCTCTCAATTTTACAATTTAAGACTTGGTG  
\*\*\*\*\*

P24 GATTGATGGGGTACAAGAACAGGTCCGAAGAAATAGGGCTGCCAATCCTCCAGTTAACAT  
P76 GATTGATGGGGTACAAGAACAGGTCCGAAGAAATAGGGCTGCCAATCCTCCAGTTAACAT  
P73 GATTGATGGGGTACAAGAACAGGTCCGAAGAAATAGGGCTGCCAATCCTCCAGTTAACAT  
P74 GATTGATGGGGTACAAGAACAGGTCCGAAGAAATAGGGCTGCCAATCCTCCAGTTAACAT  
P69 GATTGATGGGGTACAAGAACAGGTCCGAAGAAATAGGGCTGCCAATCCTCCAGTTAACAT  
K113\_gag GATTGATGGGGTACAAGAACAGGTCCGAAGAAATAGGGCTGCCAATCCTCCAGTTAACAT  
\*\*\*\*\* . \*\*\*\*\*

P24 AGATGCAGGTCAACTATTAGGAATAGGTCAAAATTTGGAGTACTATTAGTCAACAAGCATT  
P76 AGATGCAGATCAACTATTAGGAATAGGTCAAAATTTGGAGTACTATTAGTCAACAAGCATT  
P73 AGATGCAGATCAACTATTAGGAATAGGTCAAAATTTGGAGTACTATTAGTCAACAAGCATT  
P74 AGATGCAGATCAACTATTAGGAATAGGTCAAAATTTGGAGTACTATTAGTCAACAAGCATT  
P69 AGATGCAGATCAACTATTAGGAATAGGTCAAAATTTGGAGTACTATTAGTCAACAAGCATT  
K113\_gag AGATGCAGATCAACTATTAGGAATAGGTCAAAATTTGGAGTACTATTAGTCAACAAGCATT  
\*\*\*\*\* . \*\*\*\*\*

P24 AATGCAAAATGAGGCCATTGAGCAAGTTAGAGCTATCTGCCTTAGAGCCTGGGAAAAAAT  
P76 AATGCAAAATGAGGCCATTGAGCAAGTTAGAGCTATCTGCCTTAGAGCCTGGGAAAAAAT  
P73 AATGCAAAATGAGGCCATTGAGCAAGTTAGAGCTATCTGCCTTAGAGCCTGGGAAAAAAT  
P74 AATGCAAAATGAGGCCATTGAGCAAGTTAGAGCTATCTGCCTTAGAGCCTGGGAAAAAAT  
P69 AATGCAAAATGAGGCCATTGAGCAAGTTAGAGCTATCTGCCTTAGAGCCTGGGAAAAAAT  
K113\_gag AATGCAAAATGAGGCCATTGAGCAAGTTAGAGCTATCTGCCTTAGAGCCTGGGAAAAAAT  
\*\*\*\*\*

P24 CCAAGACCCAGGAAGTACCTGCCCTCATTTAATACAGTAAGACAAGGTTCAAAAGAGCC  
P76 CCAAGACCCAGGAAGTACCTGCCCTCATTTAATACAGTAAGACAAGGTTCAAAAGAGCC  
P73 CCAAGACCCAGGAAGTACCTGCCCTCATTTAATACAGTAAGACAAGGTTCAAAAGAGCC  
P74 CCAAGACCCAGGAAGTACCTGCCCTCATTTAATACAGTAAGACAAGGTTCAAAAGAGCC  
P69 CCAAGACCCAGGAAGTACCTGCCCTCATTTAATACAGTAAGACAAGGTTCAAAAGAGCC  
K113\_gag CCAAGACCCAGGAAGTACCTGCCCTCATTTAATACAGTAAGACAAGGTTCAAAAGAGCC  
\*\*\*\*\*

P24 CTATCCTGATTTTGTGGCAAGGCTCCAAGATGTTGCTCAAAAGTCAATTGCCGATGAAAA  
P76 CTATCCTGGTTTGTGGCAAGGCTCCAAGATGTTGCTCAAAAGTCAATTGCCGATGAAAA  
P73 CTATCCTGATTTTGTGGCAAGGCTCCAAGATGTTGCTCAAAAGTCAATTGCCGATGAAAA  
P74 CTATCCTGATTTTGTGGCAAGGCTCCAAGATGTTGCTCAAAAGTCAATTGCCGATGAAAA  
P69 CTATCCTGATTTTGTGGCAAGGCTCCAAGATGTTGCTCAAAAGTCAATTGCCGATGAAAA  
K113\_gag CTATCCTGATTTTGTGGCAAGGCTCCAAGATGTTGCTCAAAAGTCAATTGCCGATGAAAA  
\*\*\*\*\*.\*\*\* \*\*\*\*\*

P24 AGCCCGTAAGGTCATAGTGGAGTTAATGGCATATGAAAACGCCAATCCTGAGTGTCAATC  
P76 AGCCCGTAAGGTCATAGTGGAGTTAATGGCATATGAAAACGCCAATCCTGAGTGTCAATC  
P73 AGCCCGTAAGGTCATAGTGGAGTTAATGGCATATGAAAACGCCAATCCTGAGTGTCAATC  
P74 AGCCCGTAAGGTCATAGTGGAGTTAATGGCATATGAAAACGCCAATCCTGAGTGTCAATC  
P69 AGCCCGTAAGGTCATAGTGGAGTTAATGGCATATGAAAACGCCAATCCTGAGTGTCAATC  
K113\_gag AGCCCGTAAGGTCATAGTGGAGTTAATGGCATATGAAAACGCCAATCCTGAGTGTCAATC  
\*\*\*\*\*

P24 AGCCATTAAGCCATTAAGGAAAGGTTCCCTGCAGGATCAGATGTAATCTCAGAATATGT  
P76 AGCCATTAAGCTATTAAAGGAAAGGTTCCCTGCAGGATCAGATGTAATCTCAGAATATGT  
P73 AGCCATTAAGCCATTAAGGAAAGGTTCCCTGCAGGATCAGATGTAATCTCAGAATATGT  
P74 AGCCATTAAGCCATTAAGGAAAGGTTCCCTGCAGGATCAGATGTAATCTCAGAATATGT  
P69 AGCCATTAAGCCATTAAGGAAAGGTTCCCTGCAGGATCAGATGTAATCTCAGAATATGT  
K113\_gag AGCCATTAAGCCATTAAGGAAAGGTTCCCTGCAGGATCAGATGTAATCTCAGAATATGT  
\*\*\*\*\*

P24 AAAAGCCTGTGATGGAATGGGAGGAGCTATGCATAAAGCTATGCTTATGGCTCAAGCAAT  
P76 AAAAGCCTGTGATGGAATGGGAGGAGCTATGCATAAAGCTATGCTTATGGCTCAAGCAAT  
P73 AAAAGCCTGTGATGGAATGGGAGGAGCTATGCATAAAGCTATGCTTATGGCTCAAGCAAT  
P74 AAAAGCCTGTGATGGAATGGGAGGAGCTATGCATAAAGCTATGCTTATGGCTCAAGCAAT  
P69 AAAAGCCTGTGATGGAATGGGAGGAGCTATGCATAAAGCTATGCTTATGGCTCAAGCAAT  
K113\_gag AAAAGCCTGTGATGGAATGGGAGGAGCTATGCATAAAGCTATGCTTATGGCTCAAGCAAT  
\*\*\*\*\*

P24 AACAGGAGTTGTTTATAGGAGGACAAGTTAGAACATTTGGAGGAAAATGTTATAATTGTGG  
P76 AACAGGAGTTGTTTATAGGAGGACAAGTTAGAACATTTGGAGGAAAATGTTATAATTGTGG  
P73 AACAGGAGTTGTTTATAGGAGGACAAGTTAGAACATTTGGAGGAAAATGTTATAATTGTGG  
P74 AACAGGAGTTGTTTATAGGAGGACAAGTTAGAACATTTGGAGGAAAATGTTATAATTGTGG  
P69 AACAGGAGTTGTTTATAGGAGGACAAGTTAGAACATTTGGAGGAAAATGTTATAATTGTGG  
K113\_gag AACAGGAGTTGTTTATAGGAGGACAAGTTAGAACATTTGGAGGAAAATGTTATAATTGTGG  
\*\*\*\*\*

P24 TCAAATTGGTCACCTTAAAAAGAATTGCCAGTCTTAAACAAACAGAATATAACTATTCA  
P76 TCAAATTGGTCACCTTAAAAAGAATTGCCAGTCTTAAACAAACAGAATATAACTATTCA  
P73 TCAAATTGGTCACCTTAAAAAGAATTGCCAGTCTTAAACAAACAGAATATAACTATTCA  
P74 TCAAATTGGTCACCTTAAAAAGAATTGCCAGTCTTAAACAAACAGAATATAACTATTCA  
P69 TCAAATTGGTCACCTTAAAAAGAATTGCCAGTCTTAAACAAACAGAATATAACTATTCA  
K113\_gag TCAAATTGGTCACCTTAAAAAGAATTGCCAGTCTTAAACAAACAGAATATAACTATTCA  
\*\*\*\*\*.\*\*\*\*\*

P24 AGCAACTACAACAGGTAGAGAGCCACCTGACTTATGTCCAAGATGTAAAAAAGGAAAACA  
P76 AGCAACTACAACAGGTAGAGAGCCACCTGACTTATGTCCAAGATGTAAAAAAGGAAAACA  
P73 AGCAACTACAACAGGTAGAGAGCCACCTGACTTATGTCCAAGATGTAAAAAAGGAAAACA  
P74 AGCAACTACAACAGGTAGAGAGCCACCTGACTTATGTCCAAGATGTAAAAAAGGAAAACA  
P69 AGCAACTACAACAGGTAGAGAGCCACCTGACTTATGTCCAAGATGTAAAAAAGGAAAACA  
K113\_gag AGCAACTACAACAGGTAGAGAGCCACCTGACTTATGTCCAAGATGTAAAAAAGGAAAACA  
\*\*\*\*\*

P24 TTGGGCTAGTCAATGTCGTTCTTAAATTTGATAAAAAATGGGCAACCATTTGTCGGGAAACGA  
P76 TTGGGCTAGTCAATGTCGTTCTTAAATTTGATAAAAAATGGGCAACCATTTGTCGGGAAACGA  
P73 TTGGGCTAGTCAATGTCGTTTCTAGATTTGATAAAAAATGGGCAACCATTTGTCGGGAAACGA  
P74 TCGGGCTAGTCAATGTCGTTCTTAAATTTGATAAAAAATGGGCAACCATTTGTCGGGAAACGA  
P69 TTGGGCTAGTCAATGTCGTTCTTAAATTTGATAAAAAATGGGCAACCATTTGTCGGGAAACGA  
K113\_gag TTGGGCTAGTCAATGTCGTTCTTAAATTTGATAAAAAATGGGCAACCATTTGTCGGGAAACGA  
\* \* \* \* \*

P24 GCAAAGGGGCCAGCCTCAGGCCCCACAACAACTGGGGCATTCCTCAATTCAGCCATTTGT  
P76 GCAAAGGGGCCAGCCTCAGGCCCCACAACAACTGGGGCATTCCTCAATTCAGCCATTTGT  
P73 GCAAAGGGGCCAGCCTCAGGCCCCACAACAACTGGGGCATTCCTCAATTCAGCCATTTGT  
P74 GCAAAGGGGCCAGCCTCAGGCCCCACAACAACTGGGGCATTCCTCAATTCAGCCATTTGT  
P69 GCAAAGGGGCCAGCCTCAGGCCCCACAACAACTGGGGCATTCCTCAATTCAGCCATTTGT  
K113\_gag GCAAAGGGGCCAGCCTCAGGCCCCACAACAACTGGGGCATTCCTCAATTCAGCCATTTGT  
\* \* \* \* \*

P24 TCCTCAGAGTTTTCAGGGACAACAACCCCACTGTCCCAAGTGTTTCAGGGAATAAGCCA  
P76 TCCTCAGAGTTTTCAGGGACAACAACCCCACTGTCCCAAGTGTTTCAGGGAATAAGCCA  
P73 TCCTCAGAGTTTTCAGGGACAACAACCCCACTGTCCCAAGTGTTTCAGGGAATAAGCCA  
P74 TCCTCAGAGTTTTCAGGGACAACAACCCCACTGTCCCAAGTGTTTCAGGGAATAAGCCA  
P69 TCCTCAGAGTTTTCAGGGACAACAACCCCACTGTCCCAAGTGTTTCAGGGAATAAGCCA  
K113\_gag TCCTCAGAGTTTTCAGGGACAACAACCCCACTGTCCCAAGTGTTTCAGGGAATAAGCCA  
\* \* \* \* \*

P24 GTTACCACAATACAACAATTGTCCCCGCCACAAGCGGCAGTGCAGCAGTAGATTTATGT  
P76 GTTACCACAATACAACAATTGTCCCCGCCACAAGCGGCAGTGCAGCAGTAGATTTATGT  
P73 GTTACCACAATACAACAATTGTCCCCGCCACAAGCGGCAGTGCAGCAGTAGATTTATGT  
P74 GTTACCACAATACAACAATTGTCCCCGCCACAAGCGGCAGTGCAGCAGTAGATTTATGT  
P69 GTTACCACAATACAACAATTGTCCCCGCCACAAGCGGCAGTGCAGCAGTAGATTTATGT  
K113\_gag GTTACCACAATACAACAATTGTCCCCGCCACAAGCGGCAGTGCAGCAGTAGATTTATGT  
\* \* \* \* \*

P24 ACTATACAAGCAGTCTCTCTGCTTCCAGGGGAGCCCCACAAAAATCCCTACAGGGGTA  
P76 ACTATACAAGCAGTCTCTCTGCTTCCAGGGGAGCCCCACAAAAATCCCTACAGGGGTA  
P73 ACTATMCAAGCAGTCTCTCTGCTTCCAGGGGAGCCCCACAAAAATCCCTACAGGGGTA  
P74 ACTATACAAGCAGTCTCTCTGCTTCCAGGGGAGCCCCACAAAAATCCCTACAGGGGTA  
P69 ACTATACAAGCAGTCTCTCTGCTTCCAGGGGAGCCCCACAAAAATCCCTACAGGGGTA  
K113\_gag ACTATACAAGCAGTCTCTCTGCTTCCAGGGGAGCCCCACAAAAATCCCTACAGGTGTA  
\* \* \* \* \*

P24 TATGGCCCCCTGCCTG--  
P76 TATGGCCCCCTGCCTGAA  
P73 TATGGCCCCCTGCCTGA-  
P74 TATGGCCCCCTGCCTGA-  
P69 TATGGCCCCCTGCCTGA-  
K113\_gag TATGGCCCCCTGCCTGA-  
\* \* \* \* \*

## 8. Danksagung

Für das Ermöglichen dieser Arbeit und die Aufnahme in seine Arbeitsgruppe möchte ich mich herzlich bei Prof. Dr. Nikolaus Müller-Lantzsch bedanken. Darüber hinaus war die Geduld, die er während der Zeit, in der diese Arbeit entstand, aufgebracht hat, nicht selbstverständlich.

Ein besonderer Dank richtet sich an Frau Dr. Marlies Sauter, die mir während der Arbeit im Labor und den vielen Stunden des Schreibens geduldig zur Seite stand und eine große Hilfe war. Ohne ihre Denkanstöße und Ideen wäre die vorliegende Arbeit nicht möglich gewesen.

Frau Barbara Best, die mir vom Pipettieren über das Gießen von Agarosegelen bis zum Umgang mit Bakterien alle molekularbiologischen Schritte beigebracht hat, ebenfalls ein herzliches Danke. Sie hat einen nicht unerheblichen Teil zum Gelingen dieser Dissertation beigetragen und mir oft aus ausweglos scheinenden Situationen herausgeholfen.

Allen jetzigen und ehemaligen Mitarbeitern des Instituts für Virologie, mit denen ich zusammengearbeitet habe, sei an dieser Stelle ebenfalls gedankt. Speziell die Atmosphäre innerhalb der Arbeitsgruppe und der freundliche, kollegiale Umgang miteinander seien hier erwähnt.

Dem Pathologischen Institut in Bamberg danke ich für die gelungene Kooperation und das zur Verfügung Stellen der DNA-Proben.

Ferner bedanke ich mich bei dem Institut für Humangenetik der Universität des Saarlandes, hier v.a. PD. Dr. Jens Mayer und Dr. Thomas Stuhr. Ohne die Vorarbeiten und die Kooperation auf dem Gebiet der Polymerasekettenreaktion und Primersuche, wäre die Arbeit in diesem Umfang nicht zu realisieren gewesen.

Meinen Eltern, denen diese Arbeit gewidmet ist, möchte ich an dieser Stelle im besonderen danken, da sie mit ihrer Unterstützung ein steter Quell der Motivation und Inspiration vor allem in den Momenten sind, in denen es am schwersten ist.

Auch meiner Partnerin Claudia Schiekofer danke ich an dieser Stelle herzlich, da sie mir stets eine Stütze war und mir aus so manchem Tief herausgeholfen hat.

Zu guter Letzt danke ich dem Autor Greg Bear, ohne den ich nie auf die Idee gekommen wäre, mich intensiver mit dem Gebiet der Endogenen Retroviren zu beschäftigen. Seine Romane „*Das Darwin-Virus*“ und „*Die Darwin-Kinder*“ seien allen ans Herz gelegt, die sich über den rein wissenschaftlichen Aspekt hinaus mit diesem faszinierenden Thema beschäftigen wollen.

## 9. Lebenslauf

10.02.1981	geboren in Greifswald als Sohn von Angela Klatt (geb. Jarling) und Dr.med.habil Richard Klatt
September 1987 – Juli 1991	Besuch der Hanns-Eisler-Grundschule in Wusterhusen
September 1991 – Juli 2000	Besuch der Deutschorden-Gymnasium Bad Mergentheim
Juli 2000	Abitur
Januar 2001	Eintritt in die Bundeswehr
Februar 2001	Übernahme als Sanitätsoffizieranwärter
Oktober 2001	Immatrikulation an der Universität des Saarlandes und Beginn des Medizinstudiums
August / September 2003	Physikum (Erstes Staatsexamen)
März 2004	Aufnahme der Dissertation im Institut für Medizinische Mikrobiologie und Hygiene, Abteilung Virologie der Universität des Saarlandes
Oktober / November 2007	Zweites Staatsexamen